Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Seminar for Applied Mathematics

DNN Expression Rate Analysis of High-dimensional PDEs: Application to Option Pricing

D. Elbrächter and Ph. Grohs and A. Jentzen and Ch. Schwab

Research Report No. 2018-33
September 2018

Seminar für Angewandte Mathematik
Eidgenössische Technische Hochschule
CH-8092 Zürich
Switzerland

Funding SNF: 175699
Funding: Austrian Science Fund (FWF), project number P 30148

DNN Expression Rate Analysis of High-dimensional PDEs: Application to Option Pricing*

Dennis Elbrächter, Philipp Grohs, Arnulf Jentzen, and Christoph Schwab

September 24, 2018

Abstract

We analyze approximation rates by deep ReLU networks of a class of multi-variate solutions of Kolmogorov equations which arise in option pricing. Key technical devices are deep ReLU architectures capable of efficiently approximating tensor products. Combining this with results concerning the approximation of well behaved (i.e. fulfilling some smoothness properties) univariate functions, this provides insights into rates of deep ReLU approximation of multi-variate functions with tensor structures. We apply this in particular to the model problem given by the price of a european maximum option on a basket of d assets within the Black-Scholes model for european maximum option pricing. We prove that the solution to the d-variate option pricing problem can be approximated up to an ε-error by a deep ReLU network of size $\mathcal{O}\left(d^{2+\frac{1}{n}}\varepsilon^{-\frac{1}{n}}\right)$ where $n \in \mathbb{N}$ is abitrary (with the constant implied in $\mathcal{O}(\cdot)$ depending on n). The techniques developed in the constructive proof are of independent interest in the analysis of the expressive power of deep neural networks for solution manifolds of PDEs in high dimension.

Keywords: neural network approximation, low rank approximation, option pricing, high dimensional PDEs.
MSC2010 Classification: 41Axx, 35Kxx, 65-XX, 65D30

1 Introduction

1.1 Motivation

The development of new classification and regression algorithms based on deep neural networks – coined "Deep Learning" – revolutionized the area of artificial intelligence, machine learning, and data analysis [12]. More recently, these methods have been applied to the numerical solution of partial differential equations (PDEs for short) [25, 11, 9, 17, 15, 3, 8, 14]. In these works it has been empirically observed that deep learning-based methods work exceptionally well when used for the numerical solution of high dimensional problems arising in option pricing. The numerical experiments carried out in [3, 8, 14, 2] in particular suggest that deep learning-based methods may not suffer from the curse of dimensionality for these problems. In [24], a first theoretical result on rates of expression of infinite-variate generalized polynomial chaos expansions for solution manifolds of certain classes of parametric PDEs has been obtained.

Neural networks constitute a parametrized class of functions constructed by successive applications of affine mappings and coordinatewise nonlinearities, see [23] for a mathematical introduction. As in [22], we introduce a neural network via a tuple of matrix vector pairs

$$\Phi = (((A^1_{i,j})_{i,j=1}^{N_1,N_0}, (b^1_i)_{i=1}^{N_1}), \ldots, ((A^L_{i,j})_{i,j=1}^{N_L,N_{L-1}}, (b^L_i)_{i=1}^{N_L})) \in \times_{l=1}^{L} \left(\mathbb{R}^{N_l \times N_{l-1}} \times \mathbb{R}^{N_l}\right) \quad (1.1)$$

*This work was performed during visits of PG at the Seminar for Applied Mathematics and the FIM of ETH Zürich, and completed during the thematic term "Numerical Analysis of Complex PDE Models in the Sciences" at the Erwin Schrödinger Institute, Vienna, from June-August, 2018. AJ acknowledges support by the Swiss National Science Foundation under grant No. 175699 DE and PhG are supported in part by the Austrian Science Fund (FWF) under project number P 30148.

for given hyperparameters $L \in \mathbb{N}$, $N_0, N_1, \ldots, N_L \in \mathbb{N}$. Given an "activation function" $\varrho \in C(\mathbb{R}, \mathbb{R})$, a neural network Φ then describes a function $R_\varrho(\Phi) \in C(\mathbb{R}^{N_0}, \mathbb{R}^{N_L})$ that can be evaluated by the recursion

$$x_l = \varrho(A_l x_{l-1} + b_1), l = 1, \ldots, L-1, \quad [R_\varrho(\Phi)](x_0) = A_L x_{L-1} + b_L. \tag{1.2}$$

The number of nonzero values in the matrix vector tuples defining Φ describe the size of Φ which will be denoted by $\mathcal{M}(\Phi)$. We refer to Setting 5.1 for a more detailed description. A popular activation function ϱ is the so-called *"Rectified Linear Unit"* $\mathrm{ReLU}(x) = \max\{x, 0\}$ [12].

An increasing body of research addresses the approximation properties (or "expressive power") of deep neural networks, where by "approximation properties" we mean the study of the optimal tradeoff between the size $\mathcal{M}(\Phi)$ and the approximation error $\|u - R_\varrho(\Phi)\|$ of neural networks approximating functions u from a given function class. Classical references include [16, 7, 1, 6] as well as the summary [23] and the references therein. In these works it is shown that deep neural networks provide optimal approximation rates for classical smoothness spaces such as Sobolev spaces or Besov spaces. More recently these results have been extended to Shearlet and Ridgelet spaces [4], Modulation spaces [21], piecewise smooth functions [22] and polynomial chaos expansions [24]. All these results indicate that all classical approximation methods based on sparse expansions can be emulated by neural networks.

1.2 Contributions and Main Result

As a first main contribution of this work we show in Proposition 6.4 that low-rank functions of the form

$$(x_1, \ldots, x_d) \in \mathbb{R}^d \mapsto \sum_{s=1}^{R} c_s \prod_{j=1}^{d} h_j^s(x_j), \tag{1.3}$$

with $h_j^s \in C(\mathbb{R}, \mathbb{R})$ sufficiently regular, and $(c_s)_{s=1}^R \subseteq \mathbb{R}$ can be approximated to a given relative precision by deep ReLU neural networks of size scaling like Rd^2, that is, without curse of dimensionality. In other words, we show that in addition all classical approximation methods based on sparse expansions and on more general low-rank structures, can be emulated by neural networks. Since the solutions of several classes of high-dimensional PDEs are precisely of this form (see, e.g., [24]), our approximation results can be directly applied to these problems to establish approximation rates for neural network approximations that do not suffer from the curse of dimensionality.

As a particular application of the tools developed in the present paper, we provide a mathematical analysis of the rates of expressive power of DNNs for a particular, high-dimensional PDE which arises in mathematical finance, namely the pricing of a so-called *european maximum Option*; cf., e.g., `http://www.investment-and-finance.net/derivatives/m/maximum-option.html`).

We consider the particular (and not quite realistic) situation that the log-returns of these d assets are uncorrelated, i.e. their log-returns evolve according to d uncorrelated drifted scalar diffusion processes.

The price of the european maximum Option on this basket of d assets can then be obtained as solution of the multivariate Black-Scholes equation which reads, for the presently considered case of uncorrelated assets, as

$$(\tfrac{\partial}{\partial t} u)(t, x) + \tfrac{\mu}{2} \sum_{i=1}^{d} x_i \big(\tfrac{\partial}{\partial x_i} u\big)(t, x) + \tfrac{\sigma^2}{2} \sum_{i=1}^{d} |x_i|^2 \big(\tfrac{\partial^2}{\partial x_i^2} u\big)(t, x) = 0 \,. \tag{1.4}$$

For the european maximum option, (1.4) is completed with the *terminal condition*

$$u(T, x) = \varphi(x) = \max\{x_1 - K_1, x_2 - K_2, \ldots, x_d - K_d, 0\} \tag{1.5}$$

for $x = (x_1, \ldots, x_d) \in (0, \infty)^d$. It is well known (see, e.g., [10] and the references there) that there is a unique classical solution in $(0, \infty)^d \times [0, T]$ of the linear, parabolic equation (1.4) which attains continuously the terminal condition (1.5). This solution can be expressed as conditional expectation of the function $\varphi(x)$ in (1.5) over suitable sample paths of a d-dimensional diffusion.

One main result of this paper is the following result (stated with completely detailed assumptions below as Theorem 7.3), on expression rates of deep neural networks for the basket option price $u(0, x)$ for $x \in [a, b]^d$ for some $0 < a < b < \infty$. To render their dependence on the number d of assets in the basket explicit, we write u_d in the statement of the theorem.

Theorem 1.1. *Let $n \in \mathbb{N}$, $a \in (0,\infty)$, $b \in (a,\infty)$, $(K_i)_{i\in\mathbb{N}} \subseteq [0, K_{\max})$, and let $u_d \colon (0,\infty) \times [a,b]^d \to \mathbb{R}$, $d \in \mathbb{N}$, be the functions which satisfy for every $d \in \mathbb{N}$, $c \in (0,\infty)$, and for every $(t,x) \in [0,T] \times (0,\infty)^d$*

$$(\tfrac{\partial}{\partial t} u_d)(t,x) + \tfrac{\mu}{2} \sum_{i=1}^{d} x_i \big(\tfrac{\partial}{\partial x_i} u_d\big)(t,x) + \tfrac{\sigma^2}{2} \sum_{i=1}^{d} |x_i|^2 \big(\tfrac{\partial^2}{\partial x_i^2} u_d\big)(t,x) = 0 \tag{1.6}$$

with terminal condition at $t = T$

$$u_d(T,x) = \varphi(x) = \max\{x_1 - K_1, x_2 - K_2, \ldots, x_d - K_d, 0\}, \quad \textit{for } x \in (0,\infty)^d\,. \tag{1.7}$$

Then there exists neural networks $(\Gamma_{d,\varepsilon})_{\varepsilon\in(0,1],d\in\mathbb{N}} \in \mathfrak{N}$ which satisfy

(i) that

$$\sup_{\varepsilon\in(0,1],d\in\mathbb{N}} \left[\frac{\mathcal{M}(\Gamma_{d,\varepsilon})}{d^{2+\frac{1}{n}}\varepsilon^{-\frac{1}{n}}}\right] < \infty, \tag{1.8}$$

and

(ii) for every $\varepsilon \in (0,1]$, $d \in \mathbb{N}$ that

$$\sup_{x\in[a,b]^d} |u_d(0,x) - [R_{\mathrm{ReLU}}(\Gamma_{d,\varepsilon})](x)| \le \varepsilon. \tag{1.9}$$

Informally speaking, the previous result states that the price of a d dimensional european maximum option can, for every $n \in \mathbb{N}$, be expressed on cubes $[a,b]^d$ by deep neural networks to pointwise accuracy $\varepsilon > 0$ with network size bounded as $\mathcal{O}(d^{2+1/n}\varepsilon^{-1/n})$ for arbitrary, fixed $n \in \mathbb{N}$ and with the constant implied in $\mathcal{O}(\cdot)$ independent of d and of ε (but depending on n). In other words, the price of a european maximum option on a basket of d assets can be approximated (or "expressed") by deep ReLU networks *with spectral accuracy and without curse of dimensionality*.

The proof of this result is based on a near explicit expression for the function $u_d(0,x)$ (see Section 2). It uses this expression in conjunction with regularity estimates in Section 3 and a neural network quadrature calculus and corresponding error estimates (which is of independent interest) in Section 4 to show that the function $u_d(0,x)$ possesses an approximate low-rank representation consisting of tensor products of cumulative normal distribution functions (Lemma 4.3) to which the low-rank approximation result mentioned above can be applied.

Our results thus, for the first time, prove that neural network approximation does indeed not suffer from the curse of dimensionality and achieves spectral accuracy, in the special case of european maximum option pricing for uncorrelated assets. While we admit that this constitutes a rather special problem, the proofs in this paper develop several novel deep neural network approximation results of independent interest that can be applied to more general settings where a low rank structure is implicit in high-dimensional problems.

1.3 Outline

The structure of this article is as follows. The following Section 2 provides a derivation of the semi-explicit formula for the price of european maximum options in a standard Black-Scholes setting. This formula consists of an integral of a tensor product function. In Section 3 we develop some auxiliary regularity results for the cumulative normal distribution that are of independent interest which will be used later on. In Section 4 we show that the integral appearing in the formula of Section 2 can be efficiently approximated by numerical quadrature. Section 5 introduces some basic facts related to deep ReLU networks and Section 6 develops basic approximation results for the approximation of functions which possess a tensor product structure. Finally, in Section 7 we show our main result, namely a spectral approximation rate for the approximation of european maximum options by deep ReLU networks without curse of dimensionality. In Appendix A we collect some auxiliary proofs.

2 High-dimensional derivative pricing

In this section, we briefly review the so-called Black-Scholes ("BS equation" for short) differential equation which arises, among others, as Kolmogorov equation for multivariate geometric Brownian Motion. The significance of the BS equation stems from its role in the valuation of financial derivatives on so-called baskets of risky assets, such as stocks.

2.1 Black-Scholes-PDE

We consider the BS equation, i.e., the PDE

$$(\tfrac{\partial}{\partial t}u)(t,x) + \tfrac{\mu}{2}\sum_{i=1}^{d} x_i\big(\tfrac{\partial}{\partial x_i}u\big)(t,x) + \tfrac{\sigma^2}{2}\sum_{i=1}^{d}|x_i|^2\big(\tfrac{\partial^2}{\partial x_i^2}u\big)(t,x) = 0\,. \tag{2.1}$$

This linear, parabolic equation is, for one particular type of financial contracts (so-called "european maximum option" on a basket of d stocks whose log-returns are assumed for simplicity as mutually uncorrelated) solved for $(t,x) \in [0,T] \times (0,\infty)^d$ and is endowed with the terminal condition

$$u(T,x) = \varphi(x) = \max\{x_1 - K_1, x_2 - K_2, \ldots, x_d - K_d, 0\} \tag{2.2}$$

for $x = (x_1,\ldots,x_d) \in (0,\infty)^d$. For this and definitions of other financial contracts, we refer to e.g., `http://www.investment-and-finance.net/derivatives/m/maximum-option.html`).

2.1.1 European maximum option

Proposition 2.1. *Let $d \in \mathbb{N}$, $\mu \in \mathbb{R}$, $\sigma, T, K_1,\ldots,K_d, \xi_1,\ldots,\xi_d \in (0,\infty)$, let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space, and let $W = (W^{(1)},\ldots,W^{(d)})\colon [0,T]\times\Omega \to \mathbb{R}^d$ be a standard Brownian motion and let $u \in C([0,T]\times(0,\infty)^d)$ satisfy* (2.1) *and* (2.2)*. Then for $x = (\xi_1,\ldots,\xi_d) \in (0,\infty)^d$ it holds that*

$$\begin{aligned} u(0,x) &= \mathbb{E}\Bigg[\max_{i\in\{1,2,\ldots,d\}}\Big(\max\Big\{\exp\big(\big[\mu - \tfrac{\sigma^2}{2}\big]T + \sigma W_T^{(i)}\big)\,\xi_i - K_i, 0\Big\}\Big)\Bigg] \\ &= \int_0^\infty 1 - \Bigg[\prod_{i=1}^{d}\Bigg(\int_{-\infty}^{\frac{1}{\sigma\sqrt{T}}\left[\ln\left(\frac{y+K_i}{\xi_i}\right) - (\mu - [\sigma^2/2])T\right]} \tfrac{1}{\sqrt{2\pi}}\exp\Big(-\tfrac{r^2}{2}\Big)\,dr\Bigg)\Bigg]\,dy. \end{aligned} \tag{2.3}$$

For the proof of this Proposition, we require the following result which is in principle well-known (a proof is provided for completeness in the Appendix A.1.

Lemma 2.2 (Complementary distribution function formula)**.** *Let $\mu\colon \mathcal{B}([0,\infty)) \to [0,\infty]$ be a sigma-finite measure. Then*

$$\int_0^\infty x\,\mu(dx) = \int_0^\infty \mu([x,\infty))\,dx = \int_0^\infty \mu((x,\infty))\,dx. \tag{2.4}$$

We are now in position to provide a proof of Proposition 2.1.

Proof of Proposition 2.1. The first equality follows directly from the Feynmann-Kac formula [13, Corollary 4.17]. We proceed with a proof of the second equality. Throughout this proof let $X_i\colon \Omega \to \mathbb{R}$, $i \in \{1,2,\ldots,d\}$, be random variables which satisfy for every $i \in \{1,2,\ldots,d\}$

$$X_i = \exp\big(\big[\mu - \tfrac{\sigma^2}{2}\big]T + \sigma W_T^{(i)}\big)\,\xi_i \tag{2.5}$$

and let $Y\colon \Omega \to \mathbb{R}$ be the random variable given by

$$Y = \max\{X_1 - K_1, \ldots, X_d - K_d, 0\}. \tag{2.6}$$

Observe that for every $y \in (0,\infty)$ it holds

$$
\begin{aligned}
&\mathbb{P}(Y \geq y) \\
&= 1 - \mathbb{P}(Y < y) = 1 - \mathbb{P}\Big(\max_{i \in \{1,2,\ldots,d\}} (X_i - K_i) < y \Big) \\
&= 1 - \mathbb{P}\big(\cap_{i \in \{1,2,\ldots,d\}} \{X_i - K_i < y\} \big) = 1 - \prod_{i=1}^{d} \mathbb{P}(X_i - K_i < y) \\
&= 1 - \prod_{i=1}^{d} \mathbb{P}(X_i < y + K_i) \\
&= 1 - \prod_{i=1}^{d} \mathbb{P}\Big(\exp\big(\big[\mu - \tfrac{\sigma^2}{2}\big] T + \sigma W_T^{(i)} \big)\, \xi_i < y + K_i \Big).
\end{aligned}
\tag{2.7}
$$

Hence, we obtain that for every $y \in (0,\infty)$ it holds

$$
\begin{aligned}
\mathbb{P}(Y \geq y) &= 1 - \prod_{i=1}^{d} \mathbb{P}\Big(\exp\big(\big[\mu - \tfrac{\sigma^2}{2}\big] T + \sigma W_T^{(i)} \big) < \tfrac{y+K_i}{\xi_i} \Big) \\
&= 1 - \prod_{i=1}^{d} \mathbb{P}\Big(\sigma W_T^{(i)} < \ln\Big(\tfrac{y+K_i}{\xi_i}\Big) - \big[\mu - \tfrac{\sigma^2}{2}\big] T \Big) \\
&= 1 - \prod_{i=1}^{d} \mathbb{P}\Big(\tfrac{1}{\sqrt{T}} W_T^{(i)} < \tfrac{1}{\sigma\sqrt{T}} \Big[\ln\Big(\tfrac{y+K_i}{\xi_i}\Big) - \big[\mu - \tfrac{\sigma^2}{2}\big] T \Big] \Big).
\end{aligned}
\tag{2.8}
$$

This shows that for every $y \in (0,\infty)$ it holds

$$
\mathbb{P}(Y \geq y) = 1 - \left[\prod_{i=1}^{d} \left(\int_{-\infty}^{\frac{1}{\sigma\sqrt{T}}\left[\ln\left(\frac{y+K_i}{\xi_i}\right) - \left(\mu - [\sigma^2/2]\right)T\right]} \tfrac{1}{\sqrt{2\pi}} \exp\Big(-\tfrac{r^2}{2}\Big)\, dr \right) \right]. \tag{2.9}
$$

Combining this with Lemma 2.2 completes the proof of Proposition 2.1. ☐

With Lemma 2.2 and Proposition 2.1, we may write

$$
\begin{aligned}
&u(0,x) \\
&= \mathbb{E}\Big[\varphi\Big(\exp\Big(\big[\mu - {\sigma^2}/{2}\big]\, T + \sigma W_T^{(1)} \Big)\, x_1, \ldots, \exp\Big(\big[\mu - {\sigma^2}/{2}\big]\, T + \sigma W_T^{(d)} \Big)\, x_d \Big) \Big]
\end{aligned}
\tag{2.10}
$$

("semi-explicit" formula). Let us consider the case $\mu = \sigma^2/2$, $T = \sigma = 1$, and $K_1 = \ldots = K_d = K \in (0,\infty)$. Then for every $x = (x_1,\ldots,x_d) \in (0,\infty)^d$

$$
\begin{aligned}
&u(0,x) \\
&= \mathbb{E}\Big[\varphi\Big(e^{W_T^{(1)}} x_1, \ldots, e^{W_T^{(d)}} x_d \Big) \Big] = \mathbb{E}\Big[\varphi\Big(e^{W_1^{(1)}} x_1, \ldots, e^{W_1^{(d)}} x_d \Big) \Big] \\
&= \mathbb{E}\Big[\max\Big\{ e^{W_1^{(1)}} x_1 - K, \ldots, e^{W_1^{(d)}} x_d - K, 0 \Big\} \Big] \\
&= \int_0^\infty 1 - \left[\prod_{i=1}^{d} \int_{-\infty}^{\ln(\frac{K+c}{x_d})} \tfrac{1}{\sqrt{2\pi}} \exp\Big(-\tfrac{r^2}{2}\Big)\, dr \right] dc.
\end{aligned}
\tag{2.11}
$$

3 Regularity of the Cumulative Normal Distribution

Now that we have derived an explicit formula for the solution, we establish regularity properties of the integrand function in (2.11). This will be required in order approximate the multivariate integrals by quadratures (which are subsequently realized by neural networks) in Section 4 and to apply the neural network results from Section 6 to our problem. To this end, we analyze the derivatives of the factors in the tensor product, which essentially are concatenations of the cumulative normal distribution with the natural logarithm. As this function appears in numerous closed-form option pricing formulae (see, e.g., [18]), the (Gevrey) type regularity estimates obtained in this section are of independent interest (they may, for example, also be used in the analysis of deep network expression rates and of spectral methods for option pricing).

Lemma 3.1. *Let $f\colon (0,\infty)\to\mathbb{R}$ be the function which satisfies for every $t\in(0,\infty)$ that*

$$f(t)=\tfrac{1}{\sqrt{2\pi}}\int_{-\infty}^{\ln(t)} e^{-\frac{1}{2}r^2}\,\mathrm{d}r, \tag{3.1}$$

let $g_{n,k}\colon(0,\infty)\to\mathbb{R}$, $n,k\in\mathbb{N}_0$, be the functions which satisfy for every $n,k\in\mathbb{N}_0$, $t\in(0,\infty)$ that

$$g_{n,k}(t)=t^{-n}e^{-\frac{1}{2}[\ln(t)]^2}[\ln(t)]^k, \tag{3.2}$$

and let $(\gamma_{n,k})_{n,k\in\mathbb{Z}}\subseteq\mathbb{Z}$ be the integers which satisfy for every $n,k\in\mathbb{Z}$ that

$$\gamma_{n,k}=\begin{cases}1 & : n=1, k=0\\ -\gamma_{n-1,k-1}-(n-1)\gamma_{n-1,k}+(k+1)\gamma_{n-1,k+1} & : n>1, 0\le k<n\\ 0 & : \text{else}\end{cases}. \tag{3.3}$$

Then it holds for every $n\in\mathbb{N}$ that

(i) we have that f is n-times continuously differentiable and

(ii) we have for every $t\in(0,\infty)$ that

$$f^{(n)}(t)=\tfrac{1}{\sqrt{2\pi}}\left[\sum_{k=0}^{n-1}\gamma_{n,k}\,g_{n,k}(t)\right]. \tag{3.4}$$

Proof of Lemma 3.1. We prove (i) and (ii) by induction on $n\in\mathbb{N}$. For the base case $n=1$ note that (3.1), (3.2), (3.3), the fact that the function $\mathbb{R}\ni r\mapsto e^{-\frac{1}{2}r^2}\in(0,\infty)$ is continuous, the fundamental theorem of calculus, and the chain rule yield

(A) that f is differentiable and

(B) that for every $t\in(0,\infty)$ it holds

$$f'(t)=\tfrac{1}{\sqrt{2\pi}}\,e^{-\frac{1}{2}[\ln(t)]^2}t^{-1}=\tfrac{1}{\sqrt{2\pi}}\,g_{1,0}(t)=\tfrac{1}{\sqrt{2\pi}}\,\gamma_{1,0}\,g_{1,0}(t). \tag{3.5}$$

This establishes (i) and (ii) in the base case $n=1$. For the induction step $\mathbb{N}\ni n\to n+1\in\{2,3,4,\ldots\}$ note that for every $t\in(0,\infty)$ we have

$$\tfrac{\mathrm{d}}{\mathrm{d}t}\left[e^{-\frac{1}{2}[\ln(t)]^2}\right]=-t^{-1}e^{-\frac{1}{2}[\ln(t)]^2}\ln(t). \tag{3.6}$$

Combining this and (3.2) with the product rule establishes for every $n\in\mathbb{N}$, $k\in\{0,1,\ldots,n-1\}$, $t\in(0,\infty)$ that

$$\begin{aligned}(g_{n,k})'(t)&=\tfrac{\mathrm{d}}{\mathrm{d}t}\left[t^{-n}e^{-\frac{1}{2}[\ln(t)]^2}[\ln(t)]^k\right]\\ &=-nt^{-(n+1)}e^{-\frac{1}{2}[\ln(t)]^2}[\ln(t)]^k-t^{-(n+1)}e^{-\frac{1}{2}[\ln(t)]^2}[\ln(t)]^{k+1}\\ &\quad+t^{-(n+1)}e^{-\frac{1}{2}[\ln(t)]^2}k[\ln(t)]^{\max\{k-1,0\}}\\ &=-g_{n+1,k+1}(t)-ng_{n+1,k}(t)+kg_{n+1,\max\{k-1,0\}}(t).\end{aligned} \tag{3.7}$$

Hence, we obtain that for every $n\in\mathbb{N}$, $t\in(0,\infty)$ it holds

$$\begin{aligned}&\sum_{k=0}^{n-1}\gamma_{n,k}(g_{n,k})'(t)\\ &=\sum_{k=0}^{n-1}\left[\gamma_{n,k}\left(-g_{n+1,k+1}(t)-ng_{n+1,k}(t)+kg_{n+1,\max\{k-1,0\}}(t)\right)\right]\\ &=\sum_{k=0}^{n-1}-\gamma_{n,k}\,g_{n+1,k+1}(t)+\sum_{k=0}^{n-1}-n\gamma_{n,k}\,g_{n+1,k}(t)+\sum_{k=1}^{n-1}k\gamma_{n,k}\,g_{n+1,\max\{k-1,0\}}(t)\\ &=\sum_{k=1}^{n}-\gamma_{n,k-1}\,g_{n+1,k}(t)+\sum_{k=0}^{n-1}-n\gamma_{n,k}\,g_{n+1,k}(t)+\sum_{k=0}^{n-2}(k+1)\gamma_{n,k+1}\,g_{n+1,k}(t).\end{aligned} \tag{3.8}$$

The fact that for every $n \in \mathbb{N}$ it holds that $\gamma_{n,-1} = \gamma_{n,n} = \gamma_{n,n+1} = 0$ and (3.3) therefore ensure that for every $n \in \mathbb{N}$, $t \in (0, \infty)$ we have

$$\begin{aligned} \sum_{k=0}^{n-1} \gamma_{n,k} (g_{n,k})'(t) &= \sum_{k=0}^{n} \left[\left(-\gamma_{n,k-1} - n\gamma_{n,k} + (k+1)\gamma_{n,k+1} \right) g_{n+1,k}(t) \right] \\ &= \sum_{k=0}^{n} \gamma_{n+1,k} \, g_{n+1,k}(t). \end{aligned} \tag{3.9}$$

Induction thus establishes (i) and (ii). The proof of Lemma 3.1 is thus completed. □

Using the recursive formula from above we can now bound the derivatives of f. Note that the supremum of $f^{(n)}$ is actually attained on the interval $[e^{-4n}, 1]$ and scales with n like $e^{(cn^2)}$ for some $c \in (0, \infty)$. This can directly be seen by calulating the maximum of the $g_{n,k}$ from (3.2). For our purposes, however, it is sufficient to establish that all derivatives of f are bounded on $(0, \infty)$.

Lemma 3.2. *Let $f \colon (0, \infty) \to \mathbb{R}$ be the function which satisfies for every $t \in (0, \infty)$ that*

$$f(t) = \tfrac{1}{\sqrt{2\pi}} \int_{-\infty}^{\ln(t)} e^{-\frac{1}{2}r^2} \mathrm{d}r. \tag{3.10}$$

Then it holds for every $n \in \mathbb{N}$ that

$$\sup_{t \in (0,\infty)} \left| f^{(n)}(t) \right| \le \max\left\{ (n-1)! \, 2^{n-2}, \sup_{t \in [e^{-4n}, 1]} \left| f^{(n)}(t) \right| \right\} < \infty. \tag{3.11}$$

Proof of Lemma 3.2. Throughout this proof let $g_{n,k} \colon (0, \infty) \to \mathbb{R}$, $n, k \in \mathbb{N}_0$, be the functions such that for every $n, k \in \mathbb{N}_0$, $t \in (0, \infty)$ it holds

$$g_{n,k}(t) = t^{-n} e^{-\frac{1}{2}[\ln(t)]^2} [\ln(t)]^k \tag{3.12}$$

and let $(\gamma_{n,k})_{n,k \in \mathbb{Z}} \subseteq \mathbb{Z}$ be the integers such that for every $n, k \in \mathbb{Z}$ it holds

$$\gamma_{n,k} = \begin{cases} 1 & : n = 1, k = 0 \\ -\gamma_{n-1,k-1} - (n-1)\gamma_{n-1,k} + (k+1)\gamma_{n-1,k+1} & : n > 1, 0 \le k < n \\ 0 & : \text{else} \end{cases} \tag{3.13}$$

Then Lemma 3.1 shows for every $n \in \mathbb{N}$ that

(a) we have that f is n-times continuously differentiable and

(b) we have for every $t \in (0, \infty)$ that

$$f^{(n)}(t) = \tfrac{1}{\sqrt{2\pi}} \left[\sum_{k=0}^{n-1} \gamma_{n,k} \, g_{n,k}(t) \right]. \tag{3.14}$$

In addition, observe that for every $m \in \mathbb{N}$, $t \in (0, e^{-2m}]$ holds $\frac{1}{2}\ln(t) \le -m$. This ensures that for every $m \in \mathbb{N}$, $t \in (0, e^{-2m}] \subseteq (0, 1]$ we have

$$\left| e^{-\frac{1}{2}[\ln(t)]^2} \right| = e^{\left[\ln(t)(-\frac{1}{2}\ln(t))\right]} = \left[e^{\ln(t)} \right]^{-\frac{1}{2}\ln(t)} = t^{-\frac{1}{2}\ln(t)} = \left(\tfrac{1}{t}\right)^{\frac{1}{2}\ln(t)} \le \left(\tfrac{1}{t}\right)^{-m} = t^m. \tag{3.15}$$

Moreover, note that the fundamental theorem of calculus implies for every $t \in (0, 1]$ that

$$|\ln(t)| = |\ln(t) - \ln(1)| = |\ln(1) - \ln(t)| = \left| \int_t^1 \frac{1}{s} \, \mathrm{d}s \right| \le \left| \frac{1}{t}(1 - t) \right| \le t^{-1}. \tag{3.16}$$

Combining (3.12), (3.14), and (3.15) therefore establishes that for every $n \in \mathbb{N}$, $t \in (0, e^{-4n}) \subseteq (0, 1]$ it holds

$$\begin{aligned} \left|f^{(n)}(t)\right| &= \tfrac{1}{\sqrt{2\pi}} \left|\sum_{k=0}^{n-1} \gamma_{n,k}\, g_{n,k}(t)\right| = \tfrac{1}{\sqrt{2\pi}} \left|\sum_{k=0}^{n-1} \gamma_{n,k} t^{-n} e^{-\frac{1}{2}[\ln(t)]^2} [\ln(t)]^k\right| \\ &\leq \tfrac{1}{\sqrt{2\pi}} \left[\sum_{k=0}^{n-1} |\gamma_{n,k}|\, t^{n-k}\right] \leq \tfrac{1}{\sqrt{2\pi}} \left[\sum_{k=0}^{n-1} |\gamma_{n,k}|\right]. \end{aligned} \tag{3.17}$$

In addition, observe that the fundamental theorem of calculus ensures that for every $t \in [1, \infty)$ we have

$$|\ln(t)| = |\ln(t) - \ln(1)| = \left|\int_1^t \frac{1}{s}\, \mathrm{d}s\right| \leq |t - 1| \leq t. \tag{3.18}$$

This, (3.12), (3.14), and the fact that for every $t \in (0, \infty)$ it holds $|e^{-\frac{1}{2}[\ln(t)]^2}| \leq 1$ imply that for every $n \in \mathbb{N}$, $t \in (1, \infty)$ we have

$$\begin{aligned} \left|f^{(n)}(t)\right| &= \tfrac{1}{\sqrt{2\pi}} \left|\sum_{k=0}^{n-1} \gamma_{n,k}\, g_{n,k}(t)\right| = \tfrac{1}{\sqrt{2\pi}} \left|\sum_{k=0}^{n-1} \gamma_{n,k} t^{-n} e^{-\frac{1}{2}[\ln(t)]^2} [\ln(t)]^k\right| \\ &\leq \tfrac{1}{\sqrt{2\pi}} \left[\sum_{k=0}^{n-1} |\gamma_{n,k}|\, t^{-n} |\ln(t)|^k\right] \leq \tfrac{1}{\sqrt{2\pi}} \left[\sum_{k=0}^{n-1} |\gamma_{n,k}|\, t^{-n} t^k\right] \\ &= \tfrac{1}{\sqrt{2\pi}} \left[\sum_{k=0}^{n-1} |\gamma_{n,k}|\, t^{-n+k}\right] \leq \tfrac{1}{\sqrt{2\pi}} \left[\sum_{k=0}^{n-1} |\gamma_{n,k}|\right]. \end{aligned} \tag{3.19}$$

Moreover, observe that (a) assures that for every $n \in \mathbb{N}$ it holds that the function $f^{(n)}$ is continuous. This and the boundedness of the set $[e^{-4n}, 1]$ ensure that for every $n \in \mathbb{N}$ we have

$$\sup_{t \in [e^{-4n}, 1]} \left|f^{(n)}(t)\right| < \infty. \tag{3.20}$$

Combining this with (3.17) and (3.19) establishes that for every $n \in \mathbb{N}$ we have

$$\sup_{t \in (0,\infty)} \left|f^{(n)}(t)\right| \leq \max\left\{ \tfrac{1}{\sqrt{2\pi}} \left[\sum_{k=0}^{n-1} |\gamma_{n,k}|\right], \sup_{t \in [e^{-4n}, 1]} \left|f^{(n)}(t)\right| \right\} < \infty. \tag{3.21}$$

Furthermore, note that (3.13) implies that for every $n \in \{2, 3, 4, \ldots\}$ it holds

$$\begin{aligned} \sum_{k=0}^{n-1} |\gamma_{n,k}| &= \sum_{k=0}^{n-1} |-\gamma_{n-1,k-1} - (n-1)\gamma_{n-1,k} + (k+1)\gamma_{n-1,k+1}| \\ &\leq \left[\sum_{k=0}^{n-1} |\gamma_{n-1,k-1}|\right] + \left[\sum_{k=0}^{n-1} (n-1)\, |\gamma_{n-1,k}|\right] + \left[\sum_{k=0}^{n-1} (k+1)\, |\gamma_{n-1,k+1}|\right] \\ &= \left[\sum_{k=-1}^{n-2} |\gamma_{n-1,k}|\right] + \left[\sum_{k=0}^{n-1} (n-1)\, |\gamma_{n-1,k}|\right] + \left[\sum_{k=1}^{n} k\, |\gamma_{n-1,k}|\right]. \end{aligned} \tag{3.22}$$

Combining this with the fact that for every $n \in \{2, 3, 4, \ldots\}$, $k \in \mathbb{Z} \backslash \{0, 1, \ldots, n-2\}$ we have $\gamma_{n-1,k} = 0$ implies that for every $n \in \{2, 3, 4, \ldots\}$ it holds

$$\sum_{k=0}^{n-1} |\gamma_{n,k}| = \sum_{k=0}^{n-2} [(1 + (n-1) + k)\, |\gamma_{n-1,k}|] \leq (2n-2) \left[\sum_{k=0}^{n-2} |\gamma_{n-1,k}|\right] = 2(n-1) \left[\sum_{k=0}^{n-2} |\gamma_{n-1,k}|\right]. \tag{3.23}$$

The fact that $\gamma_{1,0} = 1$ hence implies that for every $n \in \mathbb{N}$ we have

$$\sum_{k=0}^{n-1} |\gamma_{n,k}| \leq (n-1)!\, 2^{n-1} \left[\sum_{k=0}^{0} |\gamma_{1,k}| \right] = (n-1)!\, 2^{n-1}. \tag{3.24}$$

Combining this and (3.21) ensures that for every $n \in \mathbb{N}$ it holds

$$\sup_{t \in (0,\infty)} \left| f^{(n)}(t) \right| \leq \max\left\{ \tfrac{1}{\sqrt{2\pi}} (n-1)!\, 2^{n-1}, \sup_{t \in [e^{-4n},1]} \left| f^{(n)}(t) \right| \right\} < \infty. \tag{3.25}$$

The proof of Lemma 3.2 is thus completed. □

In the following corollary we estimate the derivatives of the function $x \to f(\frac{K+c}{x})$ required to approximate this function by neural networks.

Corollary 3.3. *Let $n \in \mathbb{N}$, $K \in [0,\infty)$, $c, a \in (0,\infty)$, $b \in (a,\infty)$, let $f \colon (0,\infty) \to \mathbb{R}$ be the function which satisfies for every $t \in (0,\infty)$ that*

$$f(t) = \tfrac{1}{\sqrt{2\pi}} \int_{-\infty}^{\ln(t)} e^{-\frac{1}{2}r^2} \mathrm{d}r, \tag{3.26}$$

and let $h \colon [a,b] \to \mathbb{R}$ be the function which satisfies for every $x \in [a,b]$ that

$$h(x) = f(\tfrac{K+c}{x}). \tag{3.27}$$

Then it holds

(i) that f and h are infinitely often differentiable and

(ii) that

$$\max_{k \in \{0,1,\ldots,n\}} \sup_{x \in [a,b]} \left| h^{(k)}(x) \right| \leq n 2^{n-1} n! \left[\max_{k \in \{0,1,\ldots,n\}} \sup_{t \in [\frac{K+c}{b}, \frac{K+c}{a}]} \left| f^{(k)}(t) \right| \right] \max\{a^{-2n}, 1\} \max\{(K+c)^n, 1\}. \tag{3.28}$$

Proof of Corollary 3.3. Throughout this proof let $\alpha_{m,j} \in \mathbb{Z}$, $m, j \in \mathbb{Z}$, be the integers which satisfy that for every $m, j \in \mathbb{Z}$ it holds

$$\alpha_{m,j} = \begin{cases} -1 & : m = j = 1 \\ -(m-1+j)\alpha_{m-1,j} - \alpha_{m-1,j-1} & : m > 1,\ 1 \leq j \leq m \\ 0 & : \text{else} \end{cases}. \tag{3.29}$$

Note that Lemma 3.1 and the chain rule ensure that the functions f and h are infinitely often differentiable. Next we claim that for every $m \in \mathbb{N}$, $x \in [a,b]$ it holds

$$h^{(m)}(x) = \tfrac{\mathrm{d}^m}{\mathrm{d}x^m}\big(f(\tfrac{K+c}{x})\big) = \sum_{j=1}^{m} \alpha_{m,j} (K+c)^j x^{-(m+j)} \big(f^{(j)}\big(\tfrac{K+c}{x}\big)\big). \tag{3.30}$$

We prove (3.30) by induction on $m \in \mathbb{N}$. To prove the base case $m = 1$ we note that the chain rule ensures that for every $x \in [a,b]$ we have

$$\tfrac{\mathrm{d}}{\mathrm{d}x}\big(f(\tfrac{K+c}{x})\big) = -(K+c)x^{-2}\big(f'(\tfrac{K+c}{x})\big) = \alpha_{1,1}(K+c)x^{-2}\big(f'(\tfrac{K+c}{x})\big). \tag{3.31}$$

This establishes (3.30) in the base case $m = 1$. For the induction step $\mathbb{N} \ni m \to m+1 \in \mathbb{N}$ observe that the chain rule implies for every $m \in \mathbb{N}$, $x \in [a,b]$ that

$$\begin{aligned}
&\tfrac{\mathrm{d}}{\mathrm{d}x}\left[\sum_{j=1}^{m} \alpha_{m,j}(K+c)^j x^{-(m+j)} \left(f^{(j)}(\tfrac{K+c}{x})\right)\right] \\
&= -\left[\sum_{j=1}^{m} \alpha_{m,j}(K+c)^{j+1} x^{-(m+j+2)} \left(f^{(j+1)}(\tfrac{K+c}{x})\right)\right] - \left[\sum_{j=1}^{m} \alpha_{m,j}(K+c)^{j}(m+j) x^{-(m+j+1)} \left(f^{(j)}(\tfrac{K+c}{x})\right)\right] \\
&= -\left[\sum_{j=2}^{m+1} \alpha_{m,j-1}(K+c)^{j} x^{-(m+j+1)} \left(f^{(j)}(\tfrac{K+c}{x})\right)\right] - \left[\sum_{j=1}^{m} \alpha_{m,j}(K+c)^{j}(m+j) x^{-(m+j+1)} \left(f^{(j)}(\tfrac{K+c}{x})\right)\right] \\
&= \sum_{j=1}^{m+1} (-(m+j)\alpha_{m,j} - \alpha_{m,j-1})(K+c)^j x^{-(m+1+j)} \left(f^{(j)}(\tfrac{K+c}{x})\right).
\end{aligned} \tag{3.32}$$

Induction thus establishes (3.30). Next note that (3.29) ensures that for every $m \in \{2,3,\dots\}$ it holds

$$\begin{aligned}
\max_{j\in\{1,2,\dots,m\}} |\alpha_{m,j}| &= \max_{j\in\{1,2,\dots,m\}} |-(m-1+j)\alpha_{m-1,j} - \alpha_{m-1,j-1}| \\
&\le \left[\max_{j\in\{1,2,\dots,m-1\}} |(m-1+j)\alpha_{m-1,j}|\right] + \left[\max_{j\in\{1,2,\dots,m-1\}} |\alpha_{m-1,j}|\right] \\
&\le (2m-1)\left[\max_{j\in\{1,2,\dots,m-1\}} |\alpha_{m-1,j}|\right] \le 2m\left[\max_{j\in\{1,2,\dots,m-1\}} |\alpha_{m-1,j}|\right].
\end{aligned} \tag{3.33}$$

Induction hence proves that for every $m \in \mathbb{N}$ we have $\max_{j\in\{1,2,\dots,m\}} |\alpha_{m,j}| \le 2^{m-1}m!$. Combining this with (3.30) implies that for every $m \in \{1,2,\dots,n\}$, $x \in [a,b]$ we have

$$\begin{aligned}
\left|h^{(m)}(x)\right| &= \left|\sum_{j=1}^{m} \alpha_{m,j}(K+c)^j x^{-(m+j)}\big(f^{(j)}(\tfrac{K+c}{x})\big)\right| \\
&\le 2^{m-1}m!\left[\max_{j\in\{1,2,\dots,m\}} \sup_{t\in[\frac{K+c}{b},\frac{K+c}{a}]} \left|f^{(j)}(t)\right|\right] \max\{x^{-2m},1\}\left[\sum_{j=1}^{m}(K+c)^j\right] \\
&\le m2^{m-1}m!\left[\max_{j\in\{1,2,\dots,m\}} \sup_{t\in[\frac{K+c}{b},\frac{K+c}{a}]} \left|f^{(j)}(t)\right|\right] \max\{x^{-2m},1\}\max\{(K+c)^m,1\}.
\end{aligned} \tag{3.34}$$

Combining this with the fact that $\sup_{x\in[a,b]}|h(x)| = \sup_{t\in[\frac{K+c}{b},\frac{K+c}{a}]}|f(t)|$ establishes that it holds

$$\max_{k\in\{0,1,\dots,n\}} \sup_{x\in[a,b]} \left|h^{(k)}(x)\right| \le n2^{n-1}n!\left[\max_{k\in\{0,1,\dots,n\}} \sup_{t\in[\frac{K+c}{b},\frac{K+c}{a}]} \left|f^{(k)}(t)\right|\right] \max\{a^{-2n},1\}\max\{(K+c)^n,1\}. \tag{3.35}$$

This completes the proof of Corollary 3.3. ☐

Next we consider the derivatives of the functions $c \mapsto f(\frac{K+c}{x_i})$, $i \in \{1,2,\dots,d\}$, and their tensor product, which will be needed in order to approximate approximate the outer integral in (2.11) by composite Gaussian quadrature.

Corollary 3.4. *Let $n \in \mathbb{N}$, $K \in [0,\infty)$, $x \in (0,\infty)$, let $f\colon (0,\infty) \to \mathbb{R}$ be the function which satisfies for every $t \in (0,\infty)$ that*

$$f(t) = \tfrac{1}{\sqrt{2\pi}} \int_{-\infty}^{\ln(t)} e^{-\frac{1}{2}r^2}\mathrm{d}r, \tag{3.36}$$

and let $g\colon (0,\infty) \to \mathbb{R}$ be the function which satisfies for every $t \in (0,\infty)$ that

$$g(t) = f\big(\tfrac{K+t}{x}\big). \tag{3.37}$$

Then it holds

(i) that f and g are infinitely often differentiable and

(ii) that

$$\sup_{t\in(0,\infty)} \left|g^{(n)}(t)\right| \leq \left[\sup_{t\in(0,\infty)} \left|f^{(n)}(t)\right|\right] |x|^{-n} < \infty. \tag{3.38}$$

Proof of Corollary 3.4. Combining Lemma 3.2 with the chain rule implies that for every $t \in (0,\infty)$ it holds

$$\left|g^{(n)}(t)\right| = \left|\tfrac{\mathrm{d}^n}{\mathrm{d}t^n}\big(f(\tfrac{K+t}{x})\big)\right| = \left|f^{(n)}\big(\tfrac{K+t}{x}\big)\,\tfrac{1}{x^n}\right| \leq \left[\sup_{t\in(0,\infty)} \left|f^{(n)}(t)\right|\right] |x|^{-n} < \infty. \tag{3.39}$$

This completes the proof of Corollary 3.4. □

Lemma 3.5. *Let $d,n \in \mathbb{N}$, $a \in (0,\infty)$, $b \in (a,\infty)$, $K = (K_1,\dots,K_d) \in [0,\infty)^d$, $x = (x_1,\dots,x_d) \in [a,b]^d$, let $f\colon (0,\infty) \to \mathbb{R}$ be the function which satisfies for every $t \in (0,\infty)$ that*

$$f(t) = \tfrac{1}{\sqrt{2\pi}} \int_{-\infty}^{\ln(t)} e^{-\frac{1}{2}r^2}\mathrm{d}r, \tag{3.40}$$

and let $F\colon (0,\infty) \to \mathbb{R}$ be the function which satisfies for every $c \in (0,\infty)$ that

$$F(c) = 1 - \left[\prod_{i=1}^{d} f\Big(\tfrac{K_i+c}{x_i}\Big)\right]. \tag{3.41}$$

Then it holds

(i) that f and F are infinitely often differentiable and

(ii) that

$$\sup_{c\in(0,\infty)} \left|F^{(n)}(c)\right| \leq \left[\max_{k\in\{0,1,\dots,n\}} \sup_{t\in(0,\infty)} \left|f^{(k)}(t)\right|\right]^n d^n a^{-n} < \infty. \tag{3.42}$$

Proof of Lemma 3.5. Note that Lemma 3.1 ensures that f and F are infinitely often differentiable. Moreover, observe that (3.41) and the general Leibniz rule imply for every $c \in (0,\infty)$ that

$$\begin{aligned} F^{(n)}(c) &= -\tfrac{\mathrm{d}^n}{\mathrm{d}c^n}\left[\prod_{i=1}^{d} f\Big(\tfrac{K_i+c}{x_i}\Big)\right] \\ &= -\sum_{\substack{l_1,l_2,\dots,l_d\in\mathbb{N}_0,\\ \sum_{i=1}^d l_i = n}} \left[\binom{n}{l_1,l_2,\dots,l_d} \prod_{i=1}^{d} \Big(\tfrac{\mathrm{d}^{l_i}}{\mathrm{d}c^{l_i}}\Big[f\Big(\tfrac{K_i+c}{x_i}\Big)\Big]\Big)\right]. \end{aligned} \tag{3.43}$$

Next note that the fact that for every $r \in \mathbb{R}$ it holds that $e^{-\frac{1}{2}r^2} \geq 0$ ensures that

$$\sup_{t\in(0,\infty)} |f(t)| = \sup_{t\in(0,\infty)} \left|\tfrac{1}{\sqrt{2\pi}} \int_{-\infty}^{\ln(t)} e^{-\frac{1}{2}r^2}\mathrm{d}r\right| = \left|\tfrac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}r^2}\mathrm{d}r\right| = 1. \tag{3.44}$$

Corollary 3.4 hence establishes that for every $c \in [0,\infty)$, $l_1, \dots, l_d \in \mathbb{N}_0$ with $\sum_{i=1}^d l_i = n$ it holds

$$\begin{aligned}
\left| \prod_{i=1}^d \left(\tfrac{\mathrm{d}^{l_i}}{\mathrm{d}c^{l_i}} \left[f\left(\tfrac{K_i + c}{x_i} \right) \right] \right) \right| &\le \prod_{i=1}^d \left(\left[\sup_{t \in (0,\infty)} \left| f^{(l_i)}(t) \right| \right] |x_i|^{-l_i} \right) \\
&= \left[\prod_{i=1}^d |x_i|^{-l_i} \right] \left[\prod_{i=1}^d \left(\sup_{t \in (0,\infty)} \left| f^{(l_i)}(t) \right| \right) \right] \\
&\le \left[\prod_{i=1}^d |x_i|^{-l_i} \right] \left[\prod_{\substack{i \in \{1,2,\dots,d\}, \\ l_i > 0}} \left(\max_{k \in \{1,2,\dots,n\}} \sup_{t \in (0,\infty)} \left| f^{(k)}(t) \right| \right) \right] \\
&\le \left[\prod_{i=1}^d |x_i|^{-l_i} \right] \left[\prod_{\substack{i \in \{1,2,\dots,d\}, \\ l_i > 0}} \max\left\{ 1, \max_{k \in \{1,2,\dots,n\}} \sup_{t \in (0,\infty)} \left| f^{(k)}(t) \right| \right\} \right] \\
&\le \left[\prod_{i=1}^d |x_i|^{-l_i} \right] \left[\max\left\{ 1, \max_{k \in \{1,2,\dots,n\}} \sup_{t \in (0,\infty)} \left| f^{(k)}(t) \right| \right\} \right]^{(l_1 + \ldots + l_d)} \\
&= \left[\prod_{i=1}^d |x_i|^{-l_i} \right] \left[\max_{k \in \{0,1,\dots,n\}} \sup_{t \in (0,\infty)} \left| f^{(k)}(t) \right| \right]^n .
\end{aligned} \tag{3.45}$$

Moreover, note that the multinomial theorem ensures that

$$d^n = \left[\sum_{i=1}^d 1 \right]^n = \sum_{\substack{l_1, l_2, \dots, l_d \in \mathbb{N}_0, \\ \sum_{i=1}^d l_i = n}} \left[\binom{n}{l_1, l_2, \dots, l_d} \prod_{i=1}^d 1^{l_i} \right] = \sum_{\substack{l_1, l_2, \dots, l_d \in \mathbb{N}_0, \\ \sum_{i=1}^d l_i = n}} \left[\binom{n}{l_1, l_2, \dots, l_d} \right]. \tag{3.46}$$

Combining this with (3.43), (3.45), and the assumption that $x \in [a,b]^d$ implies that for every $c \in (0,\infty)$ we have

$$\begin{aligned}
\left| F^{(n)}(c) \right| &\le \left| \sum_{\substack{l_1, l_2, \dots, l_d \in \mathbb{N}_0, \\ \sum_{i=1}^d l_i = n}} \left[\binom{n}{l_1, l_2, \dots, l_d} \left[\prod_{i=1}^d |x_i|^{-l_i} \right] \left[\max_{k \in \{0,1,\dots,n\}} \sup_{t \in (0,\infty)} \left| f^{(k)}(t) \right| \right]^n \right] \right| \\
&\le a^{-n} \left[\max_{k \in \{0,1,\dots,n\}} \sup_{t \in (0,\infty)} \left| f^{(k)}(t) \right| \right]^n \left| \sum_{\substack{l_1, l_2, \dots, l_d \in \mathbb{N}_0, \\ \sum_{i=1}^d l_i = n}} \binom{n}{l_1, l_2, \dots, l_d} \right| \\
&= a^{-n} \left[\max_{k \in \{0,1,\dots,n\}} \sup_{t \in (0,\infty)} \left| f^{(k)}(t) \right| \right]^n d^n .
\end{aligned} \tag{3.47}$$

This completes the proof of Lemma 3.5. □

4 Quadrature

To approximate the function $x \mapsto u(0,x)$ from (2.11) by a neural network we need to evaluate, for arbitrary, given x, an expression of the form $\int_0^\infty F_x(c)\mathrm{d}c$ with F_x as defined in Lemma 4.2. We achieve this by proving in Lemma 4.2 that the functions F_x decay sufficiently fast for $c \to \infty$, and then employ numerical integration to show that the definite integral $\int_0^N F_x(c)\mathrm{d}c$ can be sufficiently well approximated by a weighted sum of $F_x(c_j)$ for suitable quadrature points $c_j \in (0,N)$. The representation of such a sum can be realized by neural networks. We show in Section 6 and 7 how the functions $x \mapsto F_x(c_j)$ for $(c_j) \in (0,N)$ can be realized efficiently due to their tensor product structure. We start by recalling an error bound for composite Gaussian quadrature which is explicit in the stepsize and quadrature order.

Lemma 4.1. *Let $n, M \in \mathbb{N}$, $N \in (0,\infty)$. Then there exist real numbers $(c_j)_{j=1}^{nM} \subseteq (0,N)$ and $(w_j)_{j=1}^{nM} \subseteq (0,\infty)$ such that for every $h \in C^{2n}([0,N],\mathbb{R})$ it holds*

$$\left| \int_0^N h(t)\,\mathrm{d}t - \sum_{j=1}^{nM} w_j h(c_j) \right| \leq \tfrac{1}{(2n)!} N^{2n+1} M^{-2n} \left[\sup_{\xi \in [0,N]} \left| h^{(2n)}(\xi) \right| \right]. \tag{4.1}$$

Proof of Lemma 4.1. Throughout this proof let $h \in C^{2n}([0,N],\mathbb{R})$ and $\alpha_k \in [0,N]$, $k \in \{0,1,\dots,M\}$, such that for every $k \in \{0,1,\dots,M\}$ it holds $\alpha_k = \frac{kN}{M}$. Observe that [19, Theorems 4.17, 6.11, and 6.12] ensure that for every $k \in \{0,1,\dots,M-1\}$ there exist $(\gamma_i^k)_{i=1}^n \subseteq (\alpha_k, \alpha_{k+1})$, $(\omega_i^k)_{i=1}^n \subseteq (0,\infty)$, and $\xi^k \in [\alpha_k, \alpha_{k+1}]$ such that

$$\int_{\alpha_k}^{\alpha_{k+1}} h(t)\,\mathrm{d}t - \sum_{i=1}^n \omega_i^k h(\gamma_i^k) = \frac{h^{(2n)}(\xi^k)}{(2n)!} \int_{\alpha_k}^{\alpha_{k+1}} \left[\prod_{i=1}^n (t-\gamma_i^k)^2 \right] \mathrm{d}t. \tag{4.2}$$

Next note that for every $k \in \{0,1,\dots,M-1\}$ it holds

$$\int_{\alpha_k}^{\alpha_{k+1}} \left[\prod_{i=1}^n (t-\gamma_i^k)^2 \right] \mathrm{d}t \leq \int_{\alpha_k}^{\alpha_{k+1}} \left[\prod_{i=1}^n (\alpha_k - \alpha_{k+1})^2 \right] \mathrm{d}t = \left[\tfrac{N}{M}\right]^{2n+1}. \tag{4.3}$$

Combining this with (4.2) yields that for every $k \in \{0,1,\dots,M\}$ we have

$$\left| \int_{\alpha_k}^{\alpha_{k+1}} h(t)\,\mathrm{d}t - \sum_{i=1}^n \omega_i^k h(\gamma_i^k) \right| \leq \frac{\left|h^{(2n)}(\xi^k)\right|}{(2n)!} \left[\tfrac{N}{M}\right]^{2n+1} \leq \tfrac{1}{(2n)!} \left[\tfrac{N}{M}\right]^{2n+1} \left[\sup_{\xi \in [0,N]} \left| h^{(2n)}(\xi) \right| \right]. \tag{4.4}$$

Hence, we obtain

$$\begin{aligned} \left| \int_0^N h(t)\,\mathrm{d}t - \sum_{k=0}^{M-1}\sum_{i=1}^n \omega_i^k h(\gamma_i^k) \right| &= \left| \sum_{k=0}^{M-1} \left[\int_{\alpha_k}^{\alpha_{k+1}} h(t)\,\mathrm{d}t - \sum_{i=1}^n \omega_i^k h(\gamma_i^k) \right] \right| \\ &\leq \sum_{k=0}^{M-1} \left(\tfrac{1}{(2n)!} \left(\tfrac{N}{M}\right)^{2n+1} \left[\sup_{\xi \in [0,N]} \left| h^{(2n)}(\xi) \right| \right] \right) \\ &= \tfrac{1}{(2n)!} N^{2n+1} M^{-2n} \left[\sup_{\xi \in [0,N]} \left| h^{(2n)}(\xi) \right| \right]. \end{aligned} \tag{4.5}$$

Let $(c_j)_{j=1}^{nM} \subseteq (0,N)$, $(w_j)_{j=1}^{nM} \subseteq (0,\infty)$ such that for every $i \in \{1,2,\dots,n\}$, $k \in \{0,1,\dots,M-1\}$ it holds

$$c_{kn+i} = \gamma_i^k \quad \text{and} \quad w_{kn+i} = \omega_i^k. \tag{4.6}$$

Next observe that

$$\left| \int_0^N h(t)\,\mathrm{d}t - \sum_{j=1}^{nM} w_j h(c_j) \right| = \left| \int_0^N h(t)\,\mathrm{d}t - \sum_{k=0}^{M-1}\sum_{i=1}^n \omega_i^k h(\gamma_i^k) \right|. \tag{4.7}$$

This completes the proof of Lemma 4.1. □

In the following we bound the error due to truncating the domain of integration.

Lemma 4.2. *Let $d, n \in \mathbb{N}$, $a \in (0,\infty)$, $b \in (a,\infty)$, $K = (K_1, K_2, \dots, K_d) \in [0,\infty)^d$, let $F_x \colon (0,\infty) \to \mathbb{R}$, $x \in [a,b]^d$, be the functions which satisfy for every $x = (x_1, x_2, \dots, x_d) \in [a,b]^d$, $c \in (0,\infty)$ that*

$$F_x(c) = 1 - \prod_{i=1}^d \left[\tfrac{1}{\sqrt{2\pi}} \int_{-\infty}^{\ln(\frac{K_i + c}{x_i})} e^{-\frac{1}{2}r^2}\,\mathrm{d}r \right], \tag{4.8}$$

and for every $\varepsilon \in (0,1]$ let $N_\varepsilon \in \mathbb{R}$ be given by $N_\varepsilon = 2e^{2(n+1)}(b+1)^{1+\frac{1}{n}} d^{\frac{1}{n}} \varepsilon^{-\frac{1}{n}}$. Then it holds for every $\varepsilon \in (0,1]$ that

$$\sup_{x \in [a,b]^d} \left| \int_{N_\varepsilon}^\infty F_x(c)\,\mathrm{d}c \right| \leq \varepsilon. \tag{4.9}$$

Proof of Lemma 4.2. Throughout this proof let $g\colon (0,\infty)\to(0,1)$ be the function given by

$$g(t) = 1 - \tfrac{1}{\sqrt{2\pi}} \int_{-\infty}^{\ln(t)} e^{-\frac{1}{2}r^2} \mathrm{d}r. \tag{4.10}$$

Note that [5, Eq.(5)] ensures that for every $y\in[0,\infty)$ we have $\frac{2}{\sqrt{\pi}}\int_y^\infty e^{-r^2}\mathrm{d}r \le e^{-y^2}$. This implies for every $t\in[1,\infty)$ that

$$0 < g(t) = 1 - \tfrac{1}{\sqrt{2\pi}} \int_{-\infty}^{\ln(t)} e^{-\frac{1}{2}r^2}\mathrm{d}r = \tfrac{1}{\sqrt{2\pi}} \int_{\ln(t)}^{\infty} e^{-\frac{1}{2}r^2}\mathrm{d}r = \tfrac{1}{\sqrt{\pi}} \int_{\frac{\ln(t)}{\sqrt{2}}}^{\infty} e^{-r^2}\mathrm{d}r \le \tfrac{1}{2} e^{-\frac{1}{2}[\ln(t)]^2}. \tag{4.11}$$

Furthermore, observe that for every $t\in[e^{2(n+1)},\infty)$ it holds

$$e^{-\frac{1}{2}[\ln(t)]^2} = e^{\left[\ln(t)(-\frac{1}{2}\ln(t))\right]} = \left[e^{\ln(t)}\right]^{-\frac{1}{2}\ln(t)} = t^{-\frac{1}{2}\ln(t)} \le t^{-(n+1)}. \tag{4.12}$$

This, (4.11), and the fact that for every $\varepsilon\in(0,1]$, $c\in[N_\varepsilon,\infty)$, $x\in[a,b]^d$, $i\in\{1,2,\dots,d\}$ we have $\frac{K_i+c}{x_i}\ge\frac{c}{b}\ge e^{2(n+1)}\ge 1$ imply that for every $\varepsilon\in(0,1]$, $c\in[N_\varepsilon,\infty)$, $x\in[a,b]^d$ it holds

$$\begin{aligned} |F_x(c)| &= \left| 1 - \prod_{i=1}^d \left[\tfrac{1}{\sqrt{2\pi}} \int_{-\infty}^{\ln(\frac{K_i+c}{x_i})} e^{-\frac{1}{2}r^2}\mathrm{d}r \right] \right| = \left| 1 - \prod_{i=1}^d \left[1 - g(\tfrac{K_i+c}{x_i}) \right] \right| \\ &\le \left| 1 - \prod_{i=1}^d \left[1 - \tfrac{1}{2} \left[\tfrac{K_i+c}{x_i}\right]^{-(n+1)} \right] \right| \le \left| 1 - \prod_{i=1}^d \left[1 - \tfrac{1}{2} \left[\tfrac{c}{b}\right]^{-(n+1)} \right] \right|. \end{aligned} \tag{4.13}$$

Combining this with the binomial theorem and the fact that for every $i\in\{1,2,\dots,d\}$ we have $\binom{d}{i}\le\frac{d^i}{i!}\le\frac{d^i}{\exp(i\ln(i)-i+1)}\le\frac{(de)^i}{i^i}$ establishes that for every $\varepsilon\in(0,1]$, $c\in[N_\varepsilon,\infty)$, $x\in[a,b]^d$ it holds

$$\begin{aligned} |F_x(c)| &\le \left| 1 - \left(1 - \tfrac{1}{2}\left[\tfrac{c}{b}\right]^{-(n+1)}\right)^d \right| = \left| 1 - \sum_{i=0}^d \left[\binom{d}{i} \left[-\tfrac{1}{2}\left[\tfrac{c}{b}\right]^{-(n+1)} \right]^i \right] \right| \\ &\le \sum_{i=1}^d \left[\binom{d}{i} \left[\tfrac{1}{2}\right]^i \left[\tfrac{b}{c}\right]^{(n+1)i} \right] \le \sum_{i=1}^d \left[\tfrac{de}{2i}\right]^i \left[\tfrac{b}{c}\right]^{(n+1)i} \\ &= \sum_{i=1}^d \left[\tfrac{e}{2i}\right]^i \left[d\left[\tfrac{b}{c}\right]^{n+1} \right]^i \le 2d\left[\tfrac{b}{c}\right]^{n+1} \left[\sum_{i=1}^d \left[d\left[\tfrac{b}{c}\right]^{n+1} \right]^{i-1} \right] \\ &= 2d\left[\tfrac{b}{c}\right]^{n+1} \left[\sum_{i=0}^{d-1} \left[d\left[\tfrac{b}{c}\right]^{n+1} \right]^{i} \right] \le 2d\left[\tfrac{b}{c}\right]^{n+1} \left[\sum_{i=0}^{\infty} \left[d\left[\tfrac{b}{c}\right]^{n+1} \right]^{i} \right]. \end{aligned} \tag{4.14}$$

This, the geometric sum formula, and the fact that for every $\varepsilon\in(0,1]$ it holds that $N_\varepsilon\ge 2bd^{\frac{1}{n}}$ imply that for every $\varepsilon\in(0,1]$, $c\in[N_\varepsilon,\infty)$, $x\in[a,b]^d$ we have

$$|F_x(c)| \le 2d\left[\tfrac{b}{c}\right]^{n+1} \left[\frac{1}{1 - d\left[\frac{b}{c}\right]^{n+1}} \right] \le 4d\left[\tfrac{b}{c}\right]^{n+1}. \tag{4.15}$$

Hence, we obtain for every $\varepsilon\in(0,1]$, $x\in[a,b]^d$ that

$$\begin{aligned} \left| \int_{N_\varepsilon}^\infty F_x(c)\,\mathrm{d}c \right| &\le 4db^{n+1} \left| \int_{N_\varepsilon}^\infty c^{-(n+1)}\mathrm{d}c \right| = 4db^{n+1}\tfrac{1}{n}(N_\varepsilon)^{-n} \\ &= \tfrac{4}{n}db^{n+1}\left[2e^{2(n+1)}(b+1)^{1+\frac{1}{n}} d^{\frac{1}{n}} \varepsilon^{-\frac{1}{n}} \right]^{-n} \\ &= \tfrac{4}{n}db^{n+1}2^{-n}e^{-(2n^2+2n)}(b+1)^{-(n+1)}d^{-1}\varepsilon \\ &= \tfrac{4}{n}2^{-n}e^{-(2n^2+n)}\left[\tfrac{b}{b+1}\right]^{n+1}\varepsilon \le \varepsilon. \end{aligned} \tag{4.16}$$

This completes the proof of Lemma 4.2. ☐

Next we combine the result above with Lemma 4.1 in order to derive the number of terms needed in order to approximate the integral by a sum to within a prescribed error bound ε.

Lemma 4.3. *Let $n \in \mathbb{N}$, $a \in (0,\infty)$, $b \in (a,\infty)$, $(K_i)_{i\in\mathbb{N}} \subseteq [0,\infty)$, let $F_x^d\colon (0,\infty) \to \mathbb{R}$, $x \in [a,b]^d$, $d \in \mathbb{N}$, be the functions which satisfy for every $d \in \mathbb{N}$, $x = (x_1, x_2, \dots, x_d) \in [a,b]^d$, $c \in (0,\infty)$ that*

$$F_x^d(c) = 1 - \prod_{i=1}^{d} \left[\tfrac{1}{\sqrt{2\pi}} \int_{-\infty}^{\ln(\frac{K_i+c}{x_i})} e^{-\frac{1}{2}r^2} \mathrm{d}r \right], \tag{4.17}$$

and for every $d \in \mathbb{N}$, $\varepsilon \in (0,1]$ let $N_{d,\varepsilon} \in \mathbb{R}$ be given by

$$N_{d,\varepsilon} = 2e^{2(n+1)}(b+1)^{1+\frac{1}{n}} d^{\frac{1}{n}} \left[\tfrac{\varepsilon}{2}\right]^{-\frac{1}{n}}. \tag{4.18}$$

Then there exist $Q_{d,\varepsilon} \in \mathbb{N}$, $c_{\varepsilon,j}^d \in (0, N_{d,\varepsilon})$, $w_{\varepsilon,j}^d \in [0,\infty)$, $j \in \{1,2,\dots,Q_{d,\varepsilon}\}$, $d \in \mathbb{N}$, $\varepsilon \in (0,1]$, such

(i) that

$$\sup_{\varepsilon\in(0,1], d\in\mathbb{N}} \left[\frac{Q_{d,\varepsilon}}{d^{1+\frac{2}{n}}\varepsilon^{-\frac{2}{n}}} \right] < \infty \tag{4.19}$$

and

(ii) that for every $d \in \mathbb{N}$, $\varepsilon \in (0,1]$ it holds $\sum_{j=1}^{Q_{d,\varepsilon}} w_{\varepsilon,j}^d = N_{d,\varepsilon}$ and

$$\sup_{x\in[a,b]^d} \left| \int_0^\infty F_x^d(c)\,\mathrm{d}c - \sum_{j=1}^{Q_{d,\varepsilon}} w_{\varepsilon,j}^d F_x^d(c_{\varepsilon,j}^d) \right| \le \varepsilon. \tag{4.20}$$

Proof of Lemma 4.3. Note that Lemma 3.5 ensures the existence of $S_m \in \mathbb{R}$, $m \in \mathbb{N}$, such that for every $d, m \in \mathbb{N}$, $x \in [a,b]^d$ it holds

$$\sup_{c\in(0,\infty)} \left| (F_x^d)^{(m)}(c) \right| \le S_m d^m. \tag{4.21}$$

Let $Q_{d,\varepsilon} \in \mathbb{R}$, $d \in \mathbb{N}$, $\varepsilon \in (0,1]$, be given by

$$Q_{d,\varepsilon} = n \left\lceil \left[\tfrac{1}{(2n)!} (N_{d,\varepsilon})^{2n+1} S_{2n} d^{2n} \tfrac{2}{\varepsilon} \right]^{\frac{1}{2n}} \right\rceil. \tag{4.22}$$

Next observe that Lemma 4.1 (with $N \leftrightarrow N_{d,\varepsilon}$ in the notation of Lemma 4.1) establishes the existence of $c_{\varepsilon,j}^d \in (0, N_{d,\varepsilon})$, $w_{\varepsilon,j}^d \in [0,\infty)$, $j \in \{1,2,\dots,Q_{d,\varepsilon}\}$, $d \in \mathbb{N}$, $\varepsilon \in (0,1]$, such that for every $d \in \mathbb{N}$, $\varepsilon \in (0,\infty)$, $x \in [a,b]^d$ we have $\sum_{j=1}^{Q_{d,\varepsilon}} w_{\varepsilon,j}^d = N_{d,\varepsilon}$ and

$$\begin{aligned} \left| \int_0^{N_{d,\varepsilon}} F_x^d(c)\mathrm{d}c - \sum_{j=1}^{Q_{d,\varepsilon}} w_{\varepsilon,j}^d F_x^d(c_{\varepsilon,j}^d) \right| &\le \tfrac{1}{(2n)!} (N_{d,\varepsilon})^{2n+1} \left[\tfrac{Q_{d,\varepsilon}}{n} \right]^{-2n} S_{2n} d^{2n} \\ &\le \tfrac{1}{(2n)!} (N_{d,\varepsilon})^{2n+1} \left[\tfrac{1}{(2n)!} (N_{d,\varepsilon})^{2n+1} S_{2n} d^{2n} \tfrac{2}{\varepsilon} \right]^{-1} S_{2n} d^{2n} = \tfrac{\varepsilon}{2}. \end{aligned} \tag{4.23}$$

Moreover, note that Lemma 4.2 (with $N_{d,\frac{\varepsilon}{2}} \leftrightarrow N_{d,\varepsilon}$ in the notation of Lemma 4.2) and (4.23) imply for every $d \in \mathbb{N}$, $\varepsilon \in (0,1]$, $x \in [a,b]^d$ that

$$\begin{aligned} &\left| \int_0^\infty F_x^d(c)\,\mathrm{d}c - \sum_{j=1}^{Q_{d,\varepsilon}} w_{\varepsilon,j}^d F_x^d(c_{\varepsilon,j}^d) \right| \\ &\le \left| \int_0^{N_{d,\varepsilon}} F_x^d(c)\,\mathrm{d}c - \sum_{j=1}^{Q_{d,\varepsilon}} w_{\varepsilon,j}^d F_x^d(c_{\varepsilon,j}^d) \right| + \left| \int_{N_{d,\varepsilon}}^\infty F_x^d(c)\,\mathrm{d}c \right| \\ &\le \tfrac{\varepsilon}{2} + \tfrac{\varepsilon}{2} = \varepsilon. \end{aligned} \tag{4.24}$$

Furthermore, we have for every $d \in \mathbb{N}$, $\varepsilon \in (0,1]$ that

$$\begin{aligned}
Q_{d,\varepsilon} &\leq n \left(1 + \left[\tfrac{1}{(2n)!}(N_{d,\varepsilon})^{2n+1} S_{2n} d^{2n} \tfrac{2}{\varepsilon}\right]^{\frac{1}{2n}}\right) \\
&= n + n \left[\tfrac{2S_{2n}}{(2n)!}\right]^{\frac{1}{2n}} d\varepsilon^{-\frac{1}{2n}} (N_{d,\varepsilon})^{1+\frac{1}{2n}} \\
&\leq n + n \left[\tfrac{2S_{2n}}{(2n)!}\right]^{\frac{1}{2n}} d\varepsilon^{-\frac{1}{2n}} \left[4e^{2(n+1)}(b+1)^{1+\frac{1}{n}} d^{\frac{1}{n}} \varepsilon^{-\frac{1}{n}}\right]^{1+\frac{1}{2n}} \\
&= n + 4n \left[\tfrac{8S_{2n}}{(2n)!}\right]^{\frac{1}{2n}} e^{2n+3+\frac{1}{n}} \left[b+1\right]^{1+\frac{3}{2n}+\frac{1}{2n^2}} d^{1+\frac{1}{n}+\frac{1}{2n^2}} \varepsilon^{-\frac{3}{2n}-\frac{1}{2n^2}} \\
&\leq nd^{1+\frac{2}{n}}\varepsilon^{-\frac{2}{n}} + 4n \left[\tfrac{8S_{2n}}{(2n)!}\right]^{\frac{1}{2n}} e^{2n+3+\frac{1}{n}} \left[b+1\right]^{1+\frac{3}{2n}+\frac{1}{2n^2}} d^{1+\frac{2}{n}} \varepsilon^{-\frac{2}{n}}.
\end{aligned} \tag{4.25}$$

This implies

$$\sup_{\varepsilon \in (0,1], d \in \mathbb{N}} \left[\frac{Q_{d,\varepsilon}}{d^{1+\frac{2}{n}}\varepsilon^{-\frac{2}{n}}}\right] \leq n + 4n \left[\tfrac{8S_{2n}}{(2n)!}\right]^{\frac{1}{2n}} e^{2n+3+\frac{1}{n}} \left[b+1\right]^{1+\frac{3}{2n}+\frac{1}{2n^2}} < \infty. \tag{4.26}$$

The proof of Lemma 4.3 is thus completed. □

5 Basic ReLU DNN Calculus

In order to talk about neural networks we will, up to some minor changes and additions, adopt the notation of P. Petersen and F. Voigtlaender from [22]. This allows us to differentiate between a neural network, defined as a structured set of weights, and its realization, which is a function on $\mathbb{R}^d$. Note that this is almost necessary in order to talk about the complexity of neural networks, since notions like depth, size or architecture do not make sense for general functions on $\mathbb{R}^d$. Even if we know that a given function 'is' a neural network, i.e. can be written a series of affine transformations and componentwise non-linearities, there are, in general, multiple non-trivially different ways to do so.
Each of these structured sets we consider does however define a unique function. This enables us to explicitly and unambiguously construct complex neural networks from simple ones, and subsequently relate the approximation capability of a given network to its complexity. Further note that since the realization of neural network is unique we can still speak of a neural network approximating a given function when its realization does so.
Specifically, a neural network will be given by its architecture, i.e. number of layers L and layer dimensions[1] $N_0, N_1, \ldots, N_L$, as well as the weights determining the affine transformations used to compute each layer from the previous one. Note that our notion of neural networks does not attach the architecture and weights to a fixed activation function, but instead considers the realization of such a neural network with respect to a given activation function. This choice is a purely technical one here, as we always consider networks with ReLU activation function.

Setting 5.1 (Neural networks)**.** *For every $L \in \mathbb{N}$, $N_0, N_1, \ldots, N_L \in \mathbb{N}$ let $\mathcal{N}_L^{N_0,N_1,\ldots,N_L}$ be the set given by*

$$\mathcal{N}_L^{N_0,N_1,\ldots,N_L} = \times_{l=1}^{L} \left(\mathbb{R}^{N_l \times N_{l-1}} \times \mathbb{R}^{N_l}\right), \tag{5.1}$$

let $\mathfrak{N}$ be the set given by

$$\mathfrak{N} = \bigcup_{\substack{L \in \mathbb{N},\\ N_0,N_1,\ldots,N_L \in \mathbb{N}}} \mathcal{N}_L^{N_0,N_1,\ldots,N_L}, \tag{5.2}$$

let $\mathcal{L}, \mathcal{M}, \mathcal{M}_l, \dim_{\text{in}}, \dim_{\text{out}} \colon \mathfrak{N} \to \mathbb{N}$, $l \in \{1, 2, \ldots, L\}$, be the functions which satisfy for every $L \in \mathbb{N}$ and every $N_0, N_1, \ldots, N_L \in \mathbb{N}$, $\Phi = (((A^1_{i,j})_{i,j=1}^{N_1,N_0}, (b^1_i)_{i=1}^{N_1}), \ldots, ((A^L_{i,j})_{i,j=1}^{N_L,N_{L-1}}, (b^L_i)_{i=1}^{N_L})) \in \mathcal{N}_L^{N_0,N_1,\ldots,N_L}$,

[1] Often phrased as input dimension N_0 and output dimension N_L with N_l, $l \in \{1, 2, \ldots, L-1\}$ many neurons in the l'th layer.

$l \in \{1, 2, \dots, L\}$ $\mathcal{L}(\Phi) = L$, $\dim_{\mathrm{in}}(\Phi) = N_0$, $\dim_{\mathrm{out}}(\Phi) = N_L$,

$$\mathcal{M}_l(\Phi) = \sum_{i=1}^{N_l} \left[\mathbb{1}_{\mathbb{R}\setminus\{0\}}(b_i^l) + \sum_{j=1}^{N_{l-1}} \mathbb{1}_{\mathbb{R}\setminus\{0\}}(A_{i,j}^l) \right], \tag{5.3}$$

and

$$\mathcal{M}(\Phi) = \sum_{l=1}^{L} \mathcal{M}_l(\Phi). \tag{5.4}$$

For every $\varrho \in C(\mathbb{R}, \mathbb{R})$ let $\varrho^* \colon \cup_{d\in\mathbb{N}} \mathbb{R}^d \to \cup_{d\in\mathbb{N}} \mathbb{R}^d$ be the function which satisfies for every $d \in \mathbb{N}$, $x = (x_1, x_2, \dots, x_d) \in \mathbb{R}^d$ that $\varrho^*(x) = (\varrho(x_1), \varrho(x_2), \dots, \varrho(x_d))$, and for every $\varrho \in \mathcal{C}(\mathbb{R}, \mathbb{R})$ denote by $R_\varrho \colon \mathfrak{N} \to \cup_{a,b\in\mathbb{N}} C(\mathbb{R}^a, \mathbb{R}^b)$ the function which satisfies for every $L \in \mathbb{N}$, $N_0, N_1, \dots, N_L \in \mathbb{N}$, $x_0 \in \mathbb{R}^{N_0}$, and $\Phi = ((A_1, b_1), (A_2, b_2), \dots, (A_L, b_L)) \in \mathcal{N}_L^{N_0, N_1, \dots, N_L}$, with $x_1 \in \mathbb{R}^{N_1}, \dots, x_{L-1} \in \mathbb{R}^{N_{L-1}}$ given by

$$x_l = \varrho^*(A_l x_{l-1} + b_l), \qquad l = 1, ..., L-1, \tag{5.5}$$

that

$$[R_\varrho(\Phi)](x_0) = A_L x_{L-1} + b_L . \tag{5.6}$$

The quantity $\mathcal{M}(\Phi)$ simply denotes the number of non-zero entries of the network Φ, which together with its depth $\mathcal{L}(\Phi)$ will be how we measure the 'size' of a given neural network Φ. One could instead consider the number of all weights, i.e. including zeroes, of a neural network. Note, however, that for any non-degenerate neural network Φ the total number of weights is bounded from above by $\mathcal{M}(\Phi)^2 + \mathcal{M}(\Phi)$. Here, the terminology "degenerate" refers to a neural network which has neurons that can be removed without changing the realization of the NN. This implies for any neural network there also exists a non-degenerate one of smaller or equal size, which has the exact same realization. Since our primary goal is to approximate d-variate functions by networks the size of which only depends polynomially on the dimension, the above means that the qualitatively same results hold regardless of which notion of 'size' is used.

We start by introducing two basic tools for constructing new neural networks from known ones and, in Lemma 5.3 and Lemma 5.4, consider how the properties of a derived network depend on its parts. The first tool will be 'concatenation' of neural networks in (5.7), which takes two networks and provides a new network whose realization is the composition of the realization of the two constituent functions. This version of concatenation only works when using the ReLU activation function, as $\varrho(x) = \max\{0, x\}$ implies $\varrho(x) - \varrho(-x) = x$. It does, however, provide us with better control on the number (and magnitude) of the weights of the resulting network.

The second tool will be the 'parallelization' of neural networks in (5.12), which will be useful when considering linear combinations or tensor products of functions which we can already approximate. While parallelization of same-depth networks (5.10) works with arbitrary activation functions, we use for the general case that any ReLU network can easily be extended (5.11) to an arbitrary depth without changing its realization.

Setting 5.2. *Assume Setting 5.1, for every $L_1, L_2 \in \mathbb{N}$, $\Phi^i = \big((A_1^i, b_1^i), (A_2^i, b_2^i), \dots, (A_{L_i}^i, b_{L_i}^i)\big) \in \mathfrak{N}$, $i \in \{1, 2\}$, with $\dim_{\mathrm{in}}(\Phi^1) = \dim_{\mathrm{out}}(\Phi^2)$ let $\Phi^1 \odot \Phi^2 \in \mathfrak{N}$ be the neural network given by*

$$\Phi^1 \odot \Phi^2 = \left((A_1^2, b_1^2), \dots, (A_{L_2-1}^2, b_{L_2-1}^2), \left(\begin{pmatrix} A_{L_2}^2 \\ -A_{L_2}^2 \end{pmatrix}, \begin{pmatrix} b_{L_2}^2 \\ -b_{L_2}^2 \end{pmatrix} \right), \left(\begin{pmatrix} A_1^1 & -A_1^1 \end{pmatrix}, b_1^1 \right), (A_2^1, b_2^1), \dots, (A_{L_1}^1, b_{L_1}^1) \right), \tag{5.7}$$

for every $d \in \mathbb{N}$, $L \in \mathbb{N} \cap [2, \infty)$ let $\Phi_{d,L}^{\mathrm{Id}} \in \mathfrak{N}$ be the neural network given by

$$\Phi_{d,L}^{\mathrm{Id}} = \left(\left(\begin{pmatrix} \mathrm{Id}_{\mathbb{R}^d} \\ -\mathrm{Id}_{\mathbb{R}^d} \end{pmatrix}, 0 \right), \underbrace{(\mathrm{Id}_{\mathbb{R}^{2d}}, 0), \dots, (\mathrm{Id}_{\mathbb{R}^{2d}}, 0)}_{L\text{-}2 \text{ times}}, \left(\begin{pmatrix} \mathrm{Id}_{\mathbb{R}^d} & -\mathrm{Id}_{\mathbb{R}^d} \end{pmatrix}, 0 \right) \right), \tag{5.8}$$

for every $d \in \mathbb{N}$ let $\Phi_{d,1}^{\mathrm{Id}} \in \mathfrak{N}$ be the neural network given by

$$\Phi_{d,1}^{\mathrm{Id}} = ((\mathrm{Id}_{\mathbb{R}^d}, 0)), \tag{5.9}$$

for every $n, L \in \mathbb{N}$, $\Phi^j = ((A_1^j, b_1^j), (A_2^j, b_2^j), \dots, (A_L^j, b_L^j)) \in \mathfrak{N}$, $j \in \{1, 2, \dots, n\}$, *let* $\mathcal{P}_s(\Phi^1, \Phi^2, \dots, \Phi^n) \in \mathfrak{N}$ *be the neural network which satisfies*

$$\mathcal{P}_s(\Phi^1, \Phi^2, \dots, \Phi^n) = \left(\left(\begin{pmatrix} A_1^1 & & & \\ & A_1^2 & & \\ & & \ddots & \\ & & & A_1^n \end{pmatrix}, \begin{pmatrix} b_1^1 \\ b_1^2 \\ \vdots \\ b_1^n \end{pmatrix} \right), \dots, \left(\begin{pmatrix} A_L^1 & & & \\ & A_L^2 & & \\ & & \ddots & \\ & & & A_L^n \end{pmatrix}, \begin{pmatrix} b_L^1 \\ b_L^2 \\ \vdots \\ b_L^n \end{pmatrix} \right) \right), \tag{5.10}$$

for every $L, d \in \mathbb{N}$, $\Phi \in \mathfrak{N}$ *with* $\mathcal{L}(\Phi) \le L$, $\dim_{\text{out}}(\Phi) = d$, *let* $\mathcal{E}_L(\Phi) \in \mathfrak{N}$ *be the neural network given by*

$$\mathcal{E}_L(\Phi) = \begin{cases} \Phi^{\text{Id}}_{d, L - \mathcal{L}(\Phi)} \odot \Phi & : \mathcal{L}(\Phi) < L \\ \Phi & : \mathcal{L}(\Phi) = L \end{cases}, \tag{5.11}$$

and for every $n, L \in \mathbb{N}$, $\Phi^j \in \mathfrak{N}$, $j \in \{1, 2, \dots, n\}$ *with* $\max_{j \in \{1,2,\dots,n\}} \mathcal{L}(\Phi^j) = L$, *let* $\mathcal{P}(\Phi^1, \Phi^2, \dots, \Phi^n) \in \mathfrak{N}$ *denote the neural network given by*

$$\mathcal{P}(\Phi^1, \Phi^2, \dots, \Phi^n) = \mathcal{P}_s(\mathcal{E}_L(\Phi^1), \mathcal{E}_L(\Phi^2), \dots, \mathcal{E}_L(\Phi^n)). \tag{5.12}$$

Lemma 5.3. *Assume Setting 5.2, let* $\Phi^1, \Phi^2 \in \mathfrak{N}$, *and let* $\varrho \colon \mathbb{R} \to \mathbb{R}$ *be the function which satisfies for every* $t \in \mathbb{R}$ *that* $\varrho(t) = \max\{0, t\}$. *Then*

(i) for every $x \in \mathbb{R}^{\dim_{\text{in}}(\Phi^2)}$ *it holds*

$$[R_\varrho(\Phi^1 \odot \Phi^2)](x) = ([R_\varrho(\Phi^1)] \circ [R_\varrho(\Phi^2)])(x) = [R_\varrho(\Phi^1)]([R_\varrho(\Phi^2)](x)), \tag{5.13}$$

(ii) $\mathcal{L}(\Phi^1 \odot \Phi^2) = \mathcal{L}(\Phi^1) + \mathcal{L}(\Phi^2)$,

(iii) $\mathcal{M}(\Phi^1 \odot \Phi^2) \le \mathcal{M}(\Phi^1) + \mathcal{M}(\Phi^2) + \mathcal{M}_1(\Phi^1) + \mathcal{M}_{\mathcal{L}(\Phi^2)}(\Phi^2) \le 2(\mathcal{M}(\Phi^1) + \mathcal{M}(\Phi^2))$,

(iv) $\mathcal{M}_1(\Phi^1 \odot \Phi^2) = \mathcal{M}_1(\Phi^2)$,

(v) $\mathcal{M}_{\mathcal{L}(\Phi^1 \odot \Phi^2)}(\Phi^1 \odot \Phi^2) = \mathcal{M}_{\mathcal{L}(\Phi^1)}(\Phi^1)$,

(vi) $\dim_{\text{in}}(\Phi^1 \odot \Phi^2) = \dim_{\text{in}}(\Phi^2)$,

(vii) $\dim_{\text{out}}(\Phi^1 \odot \Phi^2) = \dim_{\text{out}}(\Phi^1)$,

(viii) for every $d, L \in \mathbb{N}$, $x \in \mathbb{R}^d$ *it holds that* $[R_\varrho(\Phi^{\text{Id}}_{d,L})](x) = x$, *and*

(ix) for every $L \in \mathbb{N}$, $\Phi \in \mathfrak{N}$ *with* $\mathcal{L}(\Phi) \le L$, $x \in \mathbb{R}^{\dim_{\text{in}}(\Phi)}$ *it holds that* $[R_\varrho(\mathcal{E}_L(\Phi))](x) = [R_\varrho(\Phi)](x)$.

Proof of Lemma 5.3. For every $i \in \{1, 2\}$ let $L_i \in \mathbb{N}$, $N_1^i, N_2^i, \dots, N_{L_i}^i$, $(A_l^i, b_l^i) \in \mathbb{R}^{N_l^i \times N_{l-1}^i} \times \mathbb{R}^{N_l^i}$, $l \in \{1, 2, \dots, L_i\}$ such that $\Phi^i = ((A_1^i, b_1^i), \dots, (A_{L_i}^i, b_{L_i}^i))$. Furthermore, let $(A_l, b_l) \in \mathbb{R}^{N_l \times N_{l-1}} \times \mathbb{R}^{N_l}$, $l \in \{1, 2, \dots, L_1 + L_2\}$, be the matrix-vector tuples which satisfy $\Phi_1 \odot \Phi_2 = ((A_1, b_1), \dots, (A_{L_1+L_2}, b_{L_1+L_2}))$ and let $r_l \colon \mathbb{R}^{N_0} \to \mathbb{R}^{N_l}$, $l \in \{1, 2, \dots, L_1 + L_2\}$, be the functions which satisfy for every $x \in \mathbb{R}^{N_0}$ that

$$r_l(x) = \begin{cases} \varrho^*(A_1 x + b_1) & : l = 1 \\ \varrho^*(A_l r_{l-1}(x) + b_l) & : 1 < l < L_1 + L_2 \\ A_l r_{l-1}(x) + b_l & : l = L_1 + L_2 \end{cases}. \tag{5.14}$$

Observe that for every $l \in \{1, 2, \dots, L_2 - 1\}$ holds $(A_l, b_l) = (A_l^2, b_l^2)$. This implies that for every $x \in \mathbb{R}^{N_0}$ holds

$$A_{L_2}^2 r_{L_2 - 1}(x) + b_{L_2}^2 = [R_\varrho(\Phi_2)](x). \tag{5.15}$$

Combining this with (5.7) implies for every $x \in \mathbb{R}^{N_0}$ that

$$\begin{aligned} r_{L_2}(x) &= \varrho^*(A_{L_2} r_{L_2-1}(x) + b_{L_2}) = \varrho^*\left(\begin{pmatrix} A^2_{L_2} \\ -A^2_{L_2}\end{pmatrix} r_{L_2-1}(x) + \begin{pmatrix} b^2_{L_2} \\ -b^2_{L_2}\end{pmatrix}\right) \\ &= \varrho^*\left(\begin{pmatrix} A^2_{L_2} r_{l-1}(x) + b^2_{L_2} \\ -A^2_{L_2} r_{l-1}(x) - b^2_{L_2}\end{pmatrix}\right) = \begin{pmatrix} \varrho^*([R_\varrho(\Phi^2)](x)) \\ \varrho^*(-[R_\varrho(\Phi^2)](x))\end{pmatrix} \end{aligned} \tag{5.16}$$

In addition, for every $d \in \mathbb{N}$, $y = (y_1, y_2, \ldots, y_d) \in \mathbb{R}^d$ holds

$$\varrho^*(y) - \varrho^*(-y) = (\varrho(y_1) - \varrho(-y_1), \varrho(y_2) - \varrho(-y_2), \ldots, \varrho(y_d) - \varrho(-y_d)) = y. \tag{5.17}$$

This, (5.7), and (5.16) ensure that for every $x \in \mathbb{R}^{N_0}$ holds

$$\begin{aligned} r_{L_2+1}(x) &= A_{L_2+1}\begin{pmatrix} \varrho^*([R_\varrho(\Phi^2)](x)) \\ \varrho^*(-[R_\varrho(\Phi^2)](x))\end{pmatrix} + b_{L_2+1} \\ &= A^1_1 \varrho^*([R_\varrho(\Phi^2)](x)) - A^1_1 \varrho^*(-[R_\varrho(\Phi^2)](x)) + b_{L_2+1} \\ &= A^1_1 [R_\varrho(\Phi^2)](x) + b^1_1. \end{aligned} \tag{5.18}$$

Combining this with (5.14) establishes (i). Moreover, (ii)-(vii) follow directly from (5.7). Furthermore, (5.8), (5.9), and (5.17) imply (viii). Finally, (ix) follows from (5.11) and (viii). This completes the proof of Lemma 5.3. ☐

Lemma 5.4. *Assume Setting 5.2, let $\varrho\colon \mathbb{R} \to \mathbb{R}$ be the function which satisfies for every $t \in \mathbb{R}$ that $\varrho(t) = \max\{0, t\}$, let $n \in \mathbb{N}$, let $\varphi^j \in \mathfrak{N}$, $j \in \{1, 2, \ldots, n\}$, let $d_j \in \mathbb{N}$, $j \in \{1, 2, \ldots, n\}$, be given by $d_j = \dim_{\mathrm{in}}(\varphi^j)$, let $D \in \mathbb{N}$ be given by $D = \sum_{j=1}^n d_j$, and let $\Phi \in \mathfrak{N}$ be given by $\Phi = \mathcal{P}(\varphi^1, \varphi^2, \ldots, \varphi^n)$. Then*

(i) for every $x \in \mathbb{R}^D$ it holds

$$[R_\varrho(\Phi)](x) = \left([R_\varrho(\varphi^1)](x_1, \ldots, x_{d_1}), [R_\varrho(\varphi^2)](x_{d_1+1}, \ldots, x_{d_1+d_2}), \ldots, [R_\varrho(\varphi^n)](x_{D-d_n+1}, \ldots, x_D)\right), \tag{5.19}$$

(ii) $\mathcal{L}(\Phi) = \max_{j \in \{1,2,\ldots,n\}} \mathcal{L}(\varphi^j)$,

(iii) $\mathcal{M}(\Phi) \leq 2\left(\sum_{j=1}^n \mathcal{M}(\varphi^j)\right) + 4\left(\sum_{j=1}^n \dim_{\mathrm{out}}(\varphi^j)\right) \max_{j \in \{1,2,\ldots,n\}} \mathcal{L}(\varphi^j)$,

(iv) $\mathcal{M}(\Phi) = \sum_{j=1}^n \mathcal{M}(\varphi^j)$ *provided for every $j, j' \in \{1, 2, \ldots, n\}$ holds $\mathcal{L}(\varphi^j) = \mathcal{L}(\varphi^{j'})$,*

(v) $\mathcal{M}_{\mathcal{L}(\Phi)}(\Phi) \leq \sum_{j=1}^n \max\{2\dim_{\mathrm{out}}(\varphi^j), \mathcal{M}_{\mathcal{L}(\varphi^j)}(\varphi^j)\}$,

(vi) $\mathcal{M}_1(\Phi) = \sum_{j=1}^n \mathcal{M}_1(\varphi^j)$,

(vii) $\dim_{\mathrm{in}}(\Phi) = \sum_{j=1}^n \dim_{\mathrm{in}}(\varphi^j)$, *and*

(viii) $\dim_{\mathrm{out}}(\Phi) = \sum_{j=1}^n \dim_{\mathrm{out}}(\varphi^j)$.

Proof of Lemma 5.4. Observe that Lemma 5.3 implies that for every $j \in \{1, 2, \ldots, n\}$ holds

$$R_\varrho(\mathcal{E}_{\mathcal{L}(\Phi)}(\varphi^j)) = R_\varrho(\varphi^j). \tag{5.20}$$

Combining this with (5.10) and (5.12) establishes (i). Furthermore, note that that (ii), (vi), (vii), and (viii) follow directly from (5.10) and (5.12). Moreover, (5.10) demonstrates that for every $m \in \mathbb{N}$, $\psi_i \in \mathfrak{N}$, $i \in \{1, 2, \ldots, m\}$, with $\forall i, i' \in \{1, 2, \ldots, m\}\colon \mathcal{L}(\psi^i) = \mathcal{L}(\psi^{i'})$ holds

$$\mathcal{M}(\mathcal{P}_s(\psi^1, \psi^2, \ldots, \psi^m)) = \sum_{i=1}^m \mathcal{M}(\psi^i). \tag{5.21}$$

This establishes (iv). Next, observe that Lemma 5.3, (5.11), and the fact that for every $d \in$, $L \in \mathbb{N}$ holds $\mathcal{M}(\Phi^{\mathrm{Id}}_{d,L}) \leq 2dL$ imply that for every $j \in \{1, 2, \ldots, n\}$ we have

$$\begin{aligned} \mathcal{M}(\mathcal{E}_{\mathcal{L}(\Phi)}(\varphi^j)) &\leq 2\mathcal{M}(\Phi^{\mathrm{Id}}_{\dim_{\mathrm{out}}(\varphi^j),\mathcal{L}(\Phi)-\mathcal{L}(\varphi^j)}) + 2\mathcal{M}(\varphi^j) \\ &\leq 4 \dim_{\mathrm{out}}(\varphi^j)\mathcal{L}(\Phi) + 2\mathcal{M}(\varphi^j). \end{aligned} \tag{5.22}$$

Combining this with (5.21) establishes (iii). In addition, note that (5.8), (5.9), and (5.11) ensure for every $j \in \{1, 2, \ldots, n\}$ that

$$\mathcal{M}_{\mathcal{L}(\Phi)}(\mathcal{E}_{\mathcal{L}(\Phi)}(\varphi^j)) \leq \max\{2 \dim_{\mathrm{out}}(\varphi^j), \mathcal{M}_{\mathcal{L}(\varphi^j)}(\varphi^j)\}. \tag{5.23}$$

Combining this with (5.10) establishes (v). The proof of Lemma 5.4 is thus completed. □

6 Basic Expression Rate Results

Here we begin by establishing an expression rate result for a very simple function, namely $x \mapsto x^2$ on $[0, 1]$. Our approach is based on the observation by M. Telgarsky [26], that neural networks with ReLU activation function can efficiently compute high-frequent sawtooth functions, and the idea of D. Yarotsky in [27] to use this in order to approximate the function $x \mapsto x^2$ by networks computing its linear interpolations. This can then be used to derive networks capable of efficiently approximating $(x, y) \mapsto xy$, which leads to tensor products as well as polynomials and subsequently smooth function. Note that [27] uses a slighlty different notion of neural networks, where connections between non-adjacent layers are permitted. This does, however, only require a technical modification of the proof, which does not significantly change the result. Nonetheless, the respective proofs are provided in the appendix for completeness.

Lemma 6.1. *Assume Setting 5.1 and let $\varrho \colon \mathbb{R} \to \mathbb{R}$ be the ReLU activation function given by $\varrho(t) = \max\{0, t\}$. Then there exist neural networks $(\sigma_\varepsilon)_{\varepsilon \in (0,\infty)} \subseteq \mathfrak{N}$ such that for every $\varepsilon \in (0, \infty)$*

(i) $\mathcal{L}(\sigma_\varepsilon) \leq \begin{cases} \frac{1}{2} |\log_2(\varepsilon)| + 1 & : \varepsilon < 1 \\ 1 & : \varepsilon \geq 1 \end{cases}$,

(ii) $\mathcal{M}(\sigma_\varepsilon) \leq \begin{cases} 15(\frac{1}{2} |\log_2(\varepsilon)| + 1) & : \varepsilon < 1 \\ 0 & : \varepsilon \geq 1 \end{cases}$,

(iii) $\sup_{t \in [0,1]} \left|t^2 - [R_\varrho(\sigma_\varepsilon)](t)\right| \leq \varepsilon$,

(iv) $[R_\varrho(\sigma_\varepsilon)](0) = 0$.

We can now derive the following result on approximate multiplication by neural networks, by observing that $xy = 2B^2(|(x + y)/2B|^2 - |x/2B|^2 - |y/2B|^2)$ for every $B \in (0, \infty)$, $x, y \in \mathbb{R}$.

Lemma 6.2. *Assume Setting 5.1, let $B \in (0, \infty)$, and let $\varrho \colon \mathbb{R} \to \mathbb{R}$ be the ReLU activation function given by $\varrho(t) = \max\{0, t\}$. Then there exist neural networks $(\mu_\varepsilon)_{\varepsilon \in (0,\infty)} \subseteq \mathfrak{N}$ which satisfy for every $\varepsilon \in (0, \infty)$ that*

(i) $\mathcal{L}(\mu_\varepsilon) \leq \begin{cases} \frac{1}{2} \log_2(\frac{1}{\varepsilon}) + \log_2(B) + 6 & : \varepsilon < B^2 \\ 1 & : \varepsilon \geq B^2 \end{cases}$,

(ii) $\mathcal{M}(\mu_\varepsilon) \leq \begin{cases} 90 \log_2(\frac{1}{\varepsilon}) + 180 \log_2(B) + 467 & : \varepsilon < B^2 \\ 0 & : \varepsilon \geq B^2 \end{cases}$,

(iii) $\sup_{(x,y) \in [-B,B]^2} |xy - [R_\varrho(\mu_\varepsilon)](x, y)| \leq \varepsilon$,

(iv) $\mathcal{M}_1(\mu_\varepsilon) \leq 14$, $\mathcal{M}_{\mathcal{L}(\mu_\varepsilon)}(\mu_\varepsilon) = 3$, *and*

(v) for every $x \in \mathbb{R}$ it holds that $R_\varrho[\mu_\varepsilon](0, x) = R_\varrho[\mu_\varepsilon](x, 0) = 0$.

Next we extend this result to products of any number of factors by hierarchical, pairwise multiplication.

Theorem 6.3. *Assume Setting 5.1, let* $\varrho\colon \mathbb{R} \to \mathbb{R}$ *be the ReLU activation function given by* $\varrho(t) = \max\{0, t\}$, *let* $m \in \mathbb{N} \cap [2, \infty)$, *and let* $B \in [1, \infty)$. *Then there exists a constant* $C \in \mathbb{R}$ *(which is independent of* m, B*) and neural networks* $(\Pi_\varepsilon)_{\varepsilon \in (0,\infty)} \subseteq \mathfrak{N}$ *which satisfy*

(i) $\mathcal{L}(\Pi_\varepsilon) \leq C \ln(m) \left(|\ln(\varepsilon)| + m \ln(B) + \ln(m)\right)$,

(ii) $\mathcal{M}(\Pi_\varepsilon) \leq Cm \left(|\ln(\varepsilon)| + m \ln(B) + \ln(m)\right)$,

(iii) $\displaystyle\sup_{x \in [-B,B]^m} \left| \left[\prod_{j=1}^{m} x_j \right] - [R_\varrho(\Pi_\varepsilon)](x) \right| \leq \varepsilon$, *and*

(iv) $R_\varrho \left[\Pi_\varepsilon\right](x_1, x_2, \ldots, x_m) = 0$, *if there exists* $i \in \{1, 2, \ldots, m\}$ *with* $x_i = 0$.

Proof of Theorem 6.3. Throughout this proof assume Setting 5.2, let $l = \lceil \log_2 m \rceil$, and let $\theta \in \mathcal{N}_1^{1,1}$ be the neural network given by $\theta = (0, 0)$, let $(A, b) \in \mathbb{R}^{l \times m} \times \mathbb{R}^l$ be the matrix-vector tuple given by

$$A_{i,j} = \begin{cases} 1 & : i = j, j \leq m \\ 0 & : \text{else} \end{cases} \quad \text{and} \quad b_i = \begin{cases} 0 & : i \leq m \\ 1 & : i > m \end{cases}. \tag{6.1}$$

Let further $\omega \in \mathcal{N}_2^{m,2^l}$ be the neural network given by $\omega = ((A, b))$. Note that Lemma 6.2 (with B^m as B in the notation of Lemma 6.2) ensures that there exist neural networks $(\mu_\eta)_{\eta \in (0,\infty)} \subseteq \mathfrak{N}$ such that for every $\eta \in (0, [B^m]^2)$ it holds

(A) $\mathcal{L}(\mu_\eta) \leq \frac{1}{2} \log_2(\frac{1}{\eta}) + \log_2(B^m) + 6$,

(B) $\mathcal{M}(\mu_\eta) \leq 90 \log_2(\frac{1}{\eta}) + 180 \log_2(B^m) + 467$,

(C) $\displaystyle\sup_{x,y \in [-B^m, B^m]} |xy - [R_\varrho(\mu_\eta)](x, y)| \leq \eta$,

(D) $\mathcal{M}_1(\mu_\eta) \leq 14$, $\mathcal{M}_{\mathcal{L}(\mu_\eta)}(\mu_\eta) = 3$, and

(E) for every $x \in \mathbb{R}$ it holds that $R_\varrho[\mu_\eta](0, x) = R_\varrho[\mu_\eta](x, 0) = 0$.

Let $(\nu_\varepsilon)_{\varepsilon \in (0,\infty)} \subseteq \mathfrak{N}$ be the neural networks which satisfy for every $\varepsilon \in (0, \infty)$

$$\nu_\varepsilon = \mu_{m^{-2} B^{-2m} \varepsilon}. \tag{6.2}$$

Observe that (A) implies that for every $\varepsilon \in (0, B^m) \subseteq (0, m^2 B^{4m})$ it holds

$$\begin{aligned} \mathcal{L}(\nu_\varepsilon) &\leq \tfrac{1}{2} \log_2(\tfrac{1}{m^{-2} B^{-2m} \varepsilon}) + \log_2(B^m) + 6 \\ &= \tfrac{1}{2}(\log_2(\tfrac{1}{\varepsilon}) + 2 \log_2(m) + 2m \log_2(B)) + m \log_2(B) + 6 \\ &= \tfrac{1}{2} \log_2(\tfrac{1}{\varepsilon}) + 2m \log_2(B) + \log_2(m) + 6. \end{aligned} \tag{6.3}$$

In addition, note that (B) implies that for every $\varepsilon \in (0, B^m) \subseteq (0, m^2 B^{4m})$

$$\begin{aligned} \mathcal{M}(\nu_\varepsilon) &\leq 90 \log_2(\tfrac{1}{m^{-2} B^{-2m} \varepsilon}) + 180 \log_2(B^m) + 467 \\ &= 90 \log_2(\tfrac{1}{\varepsilon}) + 360m \log_2(B) + 180 \log_2(m) + 467. \end{aligned} \tag{6.4}$$

Furthermore, (C) implies that for every $\varepsilon \in (0, B^m) \subseteq (0, m^2 B^{4m})$ holds

$$\sup_{x,y \in [-B^m, B^m]} |xy - [R_\varrho(\nu_\eta)](x, y)| \leq m^{-2} B^{-2m} \varepsilon. \tag{6.5}$$

Let $\pi_{k,\varepsilon} \in \mathfrak{N}$, $\varepsilon \in (0,\infty)$, $k \in \mathbb{N}$, be the neural networks which satisfy for every $\varepsilon \in (0,\infty)$, $k \in \mathbb{N}$

$$\pi_{k,\varepsilon} = \begin{cases} \nu_\varepsilon & : k = 1 \\ \nu_\varepsilon \circ \mathcal{P}(\pi_{k-1,\varepsilon}, \pi_{k-1,\varepsilon}) & : k > 1 \end{cases} \tag{6.6}$$

and let $(\Pi_\varepsilon)_{\varepsilon \in (0,\infty)} \subseteq \mathfrak{N}$ be neural networks given by

$$\Pi_\varepsilon = \begin{cases} \pi_{l,\varepsilon} \circ \omega & : \varepsilon < B^m \\ \theta & : \varepsilon \geq B^m \end{cases}. \tag{6.7}$$

Note that for every $\varepsilon \in (B^m, \infty)$ it holds

$$\begin{aligned} \sup_{x \in [-B,B]^m} \left| \left[\prod_{j=1}^m x_j \right] - [R_\varrho(\Pi_\varepsilon)](x) \right| &= \sup_{x \in [-B,B]^m} \left| \left[\prod_{j=1}^m x_j \right] - [R_\varrho(\theta)](x) \right| \\ &= \sup_{x \in [-B,B]^m} \left| \left[\prod_{j=1}^m x_j \right] - 0 \right| = B^m \leq \varepsilon. \end{aligned} \tag{6.8}$$

We claim that for every $k \in \{1, 2, \ldots, l\}$, $\varepsilon \in (0, B^m)$ it holds

(a) that

$$\sup_{x \in [-B,B]^{(2^k)}} \left| \left[\prod_{j=1}^{2^k} x_j \right] - [R_\varrho(\pi_{k,\varepsilon})](x) \right| \leq 4^{k-1} m^{-2} B^{(2^k - 2m)} \varepsilon, \tag{6.9}$$

(b) that $\mathcal{L}(\pi_{k,\varepsilon}) \leq k\mathcal{L}(\nu_\varepsilon)$, and

(c) that $\mathcal{M}(\pi_{k,\varepsilon}) \leq (2^k - 1)\mathcal{M}(\nu_\varepsilon) + (2^{k-1} - 1)20$.

We prove (a), (b), and (c) by induction on $k \in \{1, 2, \ldots, l\}$. Observe that (6.5) and the fact that $B \in [1, \infty)$ establishes (a) for $k = 1$. Moreover, note that (6.6) establishes (b) and (c) in the base case $k = 1$.

For the induction step $\{1, 2, \ldots, l-1\} \ni k \to k+1 \in \{2, 3, \ldots, l\}$ note that Lemma 5.3, Lemma 5.4, (6.5) and (6.6) imply that for every $k \in \{1, 2, \ldots, l-1\}$, $\varepsilon \in (0, B^m)$

$$\begin{aligned} &\sup_{x \in [-B,B]^{(2^{k+1})}} \left| \left[\prod_{j=1}^{2^{k+1}} x_j \right] - [R_\varrho(\pi_{k+1,\varepsilon})](x) \right| \\ &= \sup_{x,x' \in [-B,B]^{(2^k)}} \left| \left[\prod_{j=1}^{2^k} x_j \right] \left[\prod_{j=1}^{2^k} x'_j \right] - [R_\varrho(\pi_{k+1,\varepsilon})]\left((x, x')\right) \right| \\ &= \sup_{x,x' \in [-B,B]^{(2^k)}} \left| \left[\prod_{j=1}^{2^k} x_j \right] \left[\prod_{j=1}^{2^k} x'_j \right] - [R_\varrho(\nu_\varepsilon)]\left([R_\varrho(\pi_{k,\varepsilon})](x), [R_\varrho(\pi_{k,\varepsilon})](x')\right) \right| \\ &\leq \sup_{x,x' \in [-B,B]^{(2^k)}} \left| \left[\prod_{j=1}^{2^k} x_j \right] \left[\prod_{j=1}^{2^k} x'_j \right] - \left([R_\varrho(\pi_{k,\varepsilon})](x)\right)\left([R_\varrho(\pi_{k,\varepsilon})](x')\right) \right| \\ &\quad + \sup_{x,x' \in [-B,B]^{(2^k)}} \left| \left([R_\varrho(\pi_{k,\varepsilon})](x)\right)\left([R_\varrho(\pi_{k,\varepsilon})](x')\right) - [R_\varrho(\nu_\varepsilon)]\left([R_\varrho(\pi_{k,\varepsilon})](x), [R_\varrho(\pi_{k,\varepsilon})](x')\right) \right| \\ &\leq \sup_{x,x' \in [-B,B]^{(2^k)}} \left| \left[\prod_{j=1}^{2^k} x_j \right] \left[\prod_{j=1}^{2^k} x'_j \right] - \left([R_\varrho(\pi_{k,\varepsilon})](x)\right)\left([R_\varrho(\pi_{k,\varepsilon})](x')\right) \right| + m^{-2} B^{-2m} \varepsilon. \end{aligned} \tag{6.10}$$

Next, for every $c, \delta \in (0, \infty)$, $y, z \in [-c, c]$, $\tilde{y}, \tilde{z} \in \mathbb{R}$ with $|y - \tilde{y}|, |z - \tilde{z}| \leq \delta$ it holds

$$|yz - \tilde{y}\tilde{z}| \leq 2(|y| + |z|)\delta + \delta^2 \leq 2c\delta + \delta^2. \tag{6.11}$$

Moreover, for every $k \in \{1, 2, \dots, l\}$

$$4^{k-1} \leq 4^{l-1} = 4^{\lceil \log_2 m \rceil - 1} \leq 4^{\log_2 m} = m^2. \tag{6.12}$$

The fact that $B \in [1, \infty)$ therefore ensures that for every $k \in \{1, 2, \dots, l-1\}$, $\varepsilon \in (0, B^m)$

$$\left[4^{k-1} m^{-2} B^{(2^k - 2m)} \varepsilon\right]^2 = \left[4^{k-1} m^{-2} B^{(2^{k+1} - 2m)} \varepsilon\right] \left[4^{k-1} m^{-2} B^{-2m} \varepsilon\right] \leq \left[4^{k-1} m^{-2} B^{(2^{k+1} - 2m)} \varepsilon\right]. \tag{6.13}$$

This and (6.11) imply that for every $k \in \{1, 2, \dots, l-1\}$, $\varepsilon \in (0, B^m)$, $x, x' \in [-B, B]^{(2^k)}$

$$\begin{aligned}
&\left| \left[\prod_{j=1}^{2^k} x_j\right] \left[\prod_{j=1}^{2^k} x'_j\right] - ([R_\varrho(\pi_{k,\varepsilon})](x)) ([R_\varrho(\pi_{k,\varepsilon})](x')) \right| \\
&\leq 2B^{(2^k)} 4^{k-1} m^{-2} B^{(2^k - 2m)} \varepsilon + \left[4^{k-1} m^{-2} B^{(2^k - 2m)} \varepsilon\right]^2 \\
&\leq 3 \left[4^{k-1} m^{-2} B^{(2^{k+1} - 2m)} \varepsilon\right].
\end{aligned} \tag{6.14}$$

Combining this, (6.10), and the fact that $B \in [1, \infty)$ demonstrates that for every $k \in \{1, 2, \dots, l-1\}$, $\varepsilon \in (0, B^m)$

$$\begin{aligned}
&\sup_{x \in [-B, B]^{(2^{k+1})}} \left| \left[\prod_{j=1}^{2^{k+1}} x_j\right] - [R_\varrho(\pi_{k+1,\varepsilon})](x) \right| \\
&\leq 3 \left[4^{k-1} m^{-2} B^{(2^{k+1} - 2m)} \varepsilon\right] + m^{-2} B^{-2m} \varepsilon \\
&\leq 4^k m^{-2} B^{(2^{k+1} - 2m)} \varepsilon.
\end{aligned} \tag{6.15}$$

This establishes the claim (a). Moreover, Lemma 5.3 and Lemma 5.4 imply for every $k \in \{1, 2, \dots, l-1\}$, $\varepsilon \in (0, B^m)$ with $\mathcal{L}(\pi_{k,\varepsilon}) \leq k\mathcal{L}(\nu_\varepsilon)$ holds

$$\begin{aligned}
\mathcal{L}(\pi_{k+1,\varepsilon}) &= \mathcal{L}(\nu_\varepsilon) + \max\{\mathcal{L}(\pi_{k,\varepsilon}), \mathcal{L}(\pi_{k,\varepsilon})\} \\
&\leq \mathcal{L}(\nu_\varepsilon) + k\mathcal{L}(\nu_\varepsilon) = (k+1)\mathcal{L}(\nu_\varepsilon).
\end{aligned} \tag{6.16}$$

This establishes the claim (b). Furthermore, Lemma 5.3, Lemma 5.4, (B), and (D) imply for every $k \in \{1, 2, \dots, l-1\}$, $\varepsilon \in (0, B^m)$ with $\mathcal{M}(\pi_{k,\varepsilon}) \leq (2^k - 1)\mathcal{M}(\nu_\varepsilon) + (2^{k-1} - 1)20$ holds

$$\begin{aligned}
\mathcal{M}(\pi_{k+1,\varepsilon}) &\leq \mathcal{M}(\nu_\varepsilon) + (\mathcal{M}(\pi_{k,\varepsilon}) + \mathcal{M}(\pi_{k,\varepsilon})) + \mathcal{M}_1(\nu_\varepsilon) + \mathcal{M}_{\mathcal{L}(\mathcal{P}(\pi_{k,\varepsilon}, \pi_{k,\varepsilon}))}(\mathcal{P}(\pi_{k,\varepsilon}, \pi_{k,\varepsilon})) \\
&\leq \mathcal{M}(\nu_\varepsilon) + 2\mathcal{M}(\pi_{k,\varepsilon}) + 14 + 2\mathcal{M}_{\mathcal{L}(\nu_\varepsilon)}(\nu_\varepsilon) \leq \mathcal{M}(\nu_\varepsilon) + 2\mathcal{M}(\pi_{k,\varepsilon}) + 20 \\
&\leq \mathcal{M}(\nu_\varepsilon) + 2((2^k - 1)\mathcal{M}(\nu_\varepsilon) + (2^{k-1} - 1)20) + 20 \\
&= (2^{k+1} - 1)\mathcal{M}(\nu_\varepsilon) + (2^k - 1)20.
\end{aligned} \tag{6.17}$$

This establishes the claim (c).

Combining (a) with Lemma 5.3 and (6.7) implies for every $\varepsilon \in (0, B^m)$ the bound

$$\begin{aligned}
\sup_{x \in [-B, B]^m} \left| \left[\prod_{j=1}^{m} x_j\right] - [R_\varrho(\Pi_\varepsilon)](x) \right| &\leq \sup_{x \in [-B, B]^{(2^l)}} \left| \left[\prod_{j=1}^{2^l} x_j\right] - [R_\varrho(\pi_{l,\varepsilon})](x) \right| \\
&\leq 4^{l-1} m^{-2} B^{(2^l - 2m)} \varepsilon \\
&\leq 4^{\lceil \log_2(m) \rceil - 1} m^{-2} B^{(2^{\lceil \log_2(m) \rceil} - 2m)} \varepsilon \\
&\leq 4^{\log_2(m)} m^{-2} B^{(2^{\log_2(m)+1} - 2m)} \varepsilon \\
&\leq \left[2^{\log_2(m)}\right]^2 m^{-2} B^{(2m - 2m)} \varepsilon \leq \varepsilon.
\end{aligned} \tag{6.18}$$

This and (6.8) establish that the neural networks $(\Pi_\varepsilon)_{\varepsilon\in(0,\infty)}$ satisfy (iii). Combining (b) with Lemma 5.3, (6.3), and (6.7) ensures that for every $\varepsilon \in (0, B^m)$

$$\begin{aligned}\mathcal{L}(\Pi_\varepsilon) &= \mathcal{L}(\pi_{l,\varepsilon}) + \mathcal{L}(\omega) \leq l\mathcal{L}(\nu_\varepsilon) + 1 \leq (\log_2(m)+1)\mathcal{L}(\nu_\varepsilon) + 1 \\ &\leq \log_2(m)\log_2(\tfrac{1}{\varepsilon}) + 4\log_2(m)m\log_2(B) + 2[\log_2(m)]^2 + 12\log_2(m) + 1.\end{aligned} \tag{6.19}$$

and that for every $\varepsilon \in (B^m, \infty)$ it holds $\mathcal{L}(\Pi_\varepsilon) = \mathcal{L}(\theta) = 1$. This establishes that the neural networks $(\Pi_\varepsilon)_{\varepsilon\in(0,\infty)}$ satisfy (i). Furthermore, note that (c), Lemma 5.3, (6.3), and (6.7) demonstrate that for every $\varepsilon \in (0, B^m)$

$$\begin{aligned}\mathcal{M}(\Pi_\varepsilon) &\leq 2(\mathcal{M}(\pi_{l,\varepsilon}) + \mathcal{M}(\omega)) \leq 2\left[(2^l-1)\mathcal{M}(\nu_\varepsilon) + (2^{l-1}-1)20\right] + 4m \\ &\leq 2^{l+1}\mathcal{M}(\nu_\varepsilon) + (2^l)20 + 4m \leq 4m\mathcal{M}(\nu_\varepsilon) + 44m \\ &\leq 360m\log_2(\tfrac{1}{\varepsilon}) + 1440m^2\log_2(B) + 720m\log_2(m) + 1912m.\end{aligned} \tag{6.20}$$

and that for every $\varepsilon \in (B^m, \infty)$ holds $\mathcal{M}(\Pi_\varepsilon) = \mathcal{M}(\theta) = 0$. This establishes that the neural networks $(\Pi_\varepsilon)_{\varepsilon\in(0,\infty)}$ satisfy (ii). Note that (iv) follows from (E) by construction. The proof of Theorem 6.3 is thus completed. ☐

With the above established, it is quite straightforward to get the following result for the approximation of tensor products. Note that the exponential term B^{m-1} in (iii) is unavoidable as result from multiplying m many inaccurate values of magnitude B. For our purposes this will not be an issue since the functions we consider are bounded in absolute value by $B = 1$. This is further not an issue in cases, where the h_j can be approximated by networks whose size scales logarithmically with ε.

Proposition 6.4. *Assume Setting 5.2, let $\varrho\colon \mathbb{R} \to \mathbb{R}$ be the ReLU activation function given by $\varrho(t) = \max\{0, t\}$, let $B \in [1, \infty)$, $m \in \mathbb{N}$, for every $j \in \{1, 2, \dots, m\}$ let $d_j \in \mathbb{N}$, $\Omega_j \subseteq \mathbb{R}^{d_j}$, and $h_j : \Omega_j \to [-B, B]$, let $(\Phi^j_\varepsilon)_{\varepsilon\in(0,\infty)} \in \mathfrak{N}$, $j \in \{1, 2, \dots, m\}$, be neural networks which satisfy for every $\varepsilon \in (0, \infty)$, $j \in \{1, 2, \dots, m\}$*

$$\sup_{t\in\Omega_j} \left|h_j(x) - \left[R_\varrho(\Phi^j_\varepsilon)\right](x)\right| \leq \varepsilon, \tag{6.21}$$

let $\Phi^{\mathcal{P}}_\varepsilon \in \mathfrak{N}$, $\varepsilon \in (0, \infty)$ be be given by $\Phi^{\mathcal{P}}_\varepsilon = \mathcal{P}(\Phi^1_\varepsilon, \Phi^2_\varepsilon, \dots, \Phi^m_\varepsilon)$, and let $L_\varepsilon \in \mathbb{N}$, $\varepsilon \in (0, \infty)$ be given by $L_\varepsilon = \max_{j\in\{1,2,\dots,m\}} \mathcal{L}(\Phi^j_\varepsilon)$.
Then there exists a constant $C \in \mathbb{R}$ (which is independent of m, B, ε) and neural networks $(\Psi_\varepsilon)_{\varepsilon\in(0,\infty)} \subseteq \mathfrak{N}$ which satsify

(i) $\mathcal{L}(\Psi_\varepsilon) \leq C\ln(m)\left(|\ln(\varepsilon)| + m\ln(B) + \ln(m)\right) + L_\varepsilon$,

(ii) $\mathcal{M}(\Psi_\varepsilon) \leq Cm\left(|\ln(\varepsilon)| + m\ln(B) + \ln(m)\right) + \mathcal{M}(\Phi^{\mathcal{P}}_\varepsilon) + \mathcal{M}_{L_\varepsilon}(\Phi^{\mathcal{P}}_\varepsilon)$, *and*

(iii) $\displaystyle\sup_{t=(t_1,t_2,\dots,t_m)\in\times_{j=1}^m\Omega_j} \left|\left[\prod_{j=1}^m h_j(t_j)\right] - [R_\varrho(\Psi_\varepsilon)](t)\right| \leq 3mB^{m-1}\varepsilon.$

Proof of Proposition 6.4. In the case of $m = 1$ the neural networks $(\Phi^1_\varepsilon)_{\varepsilon\in(0,\infty)} \in \mathfrak{N}$ satisfy (i), (ii), and (iii) by assumption. Throughout the remainder of this proof assume $m \geq 2$, and let $\theta \in \mathcal{N}^{1,1}_1$ denote the trivial neural network $\theta = (0, 0)$. Observe that Theorem 6.3 (with $\varepsilon \leftrightarrow \eta$, $C' \leftrightarrow C$ in the notation Theorem 6.3) ensures that there exist $C' \in \mathbb{R}$ and neural networks $(\Pi_\eta)_{\eta\in(0,\infty)} \subseteq \mathfrak{N}$ which satisfy for every $\eta \in (0, \infty)$ that

(a) $\mathcal{L}(\Pi_\eta) \leq C'\ln(m)\left(|\ln(\eta)| + m\ln(B) + \ln(m)\right)$,

(b) $\mathcal{M}(\Pi_\eta) \leq C'm\left(|\ln(\eta)| + m\ln(B) + \ln(m)\right)$, and

(c) $\displaystyle\sup_{x\in[-B,B]^m} \left|\left[\prod_{j=1}^m x_j\right] - [R_\varrho(\Pi_\eta)](x)\right| \leq \eta.$

Let $(\Psi_\varepsilon)_{\varepsilon\in(0,\infty)} \subseteq \mathfrak{N}$ be the neural networks which satisfy for every $\varepsilon \in (0,\infty)$ that

$$\Psi_\varepsilon = \begin{cases} \Pi_\varepsilon \circ \mathcal{P}(\Phi_\varepsilon^1, \Phi_\varepsilon^2, \ldots, \Phi_\varepsilon^m) & : \varepsilon < \frac{B}{2m} \\ \theta & : \varepsilon \geq \frac{B}{2m} \end{cases}. \tag{6.22}$$

Note that for every $\varepsilon \in (0, \frac{B}{2m})$

$$\begin{aligned} \max_{\substack{x\in[-B,B]^m, x'\in\mathbb{R}^m \\ \|x'-x\|_\infty \leq \varepsilon}} \left| \prod_{j=1}^m x_j' - \prod_{j=1}^m x_j \right| &= (B+\varepsilon)^m - B^m = \sum_{k=1}^m \binom{m}{k} B^{m-k}\varepsilon^k \leq \varepsilon \sum_{k=1}^m \frac{m^k}{k!} B^{m-k}\varepsilon^{k-1} \\ &\leq \varepsilon \sum_{k=1}^m \frac{m^k}{k!} B^{m-k} \left(\frac{B}{2m}\right)^{k-1} = mB^{m-1}\varepsilon \sum_{k=1}^m \frac{1}{2^{k-1}k!} \\ &\leq 2mB^{m-1}\varepsilon. \end{aligned} \tag{6.23}$$

Combining this with Lemma 5.3, Lemma 5.4, (6.21), and (c) implies that for every $\varepsilon \in (0, \frac{B}{2m})$, $t = (t_1, t_2, \ldots, t_m) \in \Omega$ it holds

$$\begin{aligned} \left| \left[\prod_{j=1}^m h_j(t_j)\right] - [R_\varrho(\Psi_\varepsilon)](t) \right| &= \left| \left[\prod_{j=1}^m h_j(t_j)\right] - \left[R_\varrho(\Pi_\varepsilon \circ \mathcal{P}(\Phi_\varepsilon^1, \Phi_\varepsilon^2, \ldots, \Phi_\varepsilon^m))\right](t) \right| \\ &\leq \left| \left[\prod_{j=1}^m h_j(t_j)\right] - \left[\prod_{j=1}^m \left[R_\varrho(\Phi_\varepsilon^j)\right](t_j)\right] \right| \\ &\quad + \left| \left[\prod_{j=1}^m \left[R_\varrho(\Phi_\varepsilon^j)\right](t_j)\right] - [R_\varrho(\Pi_\varepsilon)]\left([R_\varrho(\Phi_\varepsilon^1)](t_1), \ldots, [R_\varrho(\Phi_\varepsilon^m)](t_j)\right) \right| \\ &\leq 2mB^{m-1}\varepsilon + \varepsilon \leq 3mB^{m-1}\varepsilon. \end{aligned} \tag{6.24}$$

Moreover, for every $\varepsilon \in [\frac{B}{2m}, \infty)$, $t = (t_1, t_2, \ldots, t_m) \in \Omega$ it holds that

$$\begin{aligned} \left| \left[\prod_{j=1}^m h_j(t_j)\right] - [R_\varrho(\Psi_\varepsilon)](t) \right| &= \left| \left[\prod_{j=1}^m h_j(t_j)\right] - [R_\varrho(\theta)](t) \right| \\ &= \left| \left[\prod_{j=1}^m h_j(t_j)\right] \right| \leq B^m \leq 2mB^{m-1}\varepsilon. \end{aligned} \tag{6.25}$$

This and (6.24) establish that the neural networks $(\Psi_\varepsilon)_{\varepsilon,c\in(0,\infty)}$ satisfy (iii). Next observe that Lemma 5.3, Lemma 5.4, and (a) demonstrate that for every $\varepsilon \in (0, \frac{B}{2m})$

$$\begin{aligned} \mathcal{L}(\Psi_\varepsilon) &= \mathcal{L}(\Pi_\varepsilon \circ \mathcal{P}(\Phi_\varepsilon^1, \Phi_\varepsilon^2, \ldots, \Phi_\varepsilon^m)) = \mathcal{L}(\Pi_\varepsilon) + \max_{j\in\{1,2,\ldots,m\}} \mathcal{L}(\Phi_\varepsilon^j) \\ &\leq C' \ln(m) \left(|\ln(\varepsilon)| + m\ln(B) + \ln(m)\right) + L_\varepsilon. \end{aligned} \tag{6.26}$$

This and the fact that for every $\varepsilon \in [\frac{B}{2m}, \infty)$ it holds that $\mathcal{L}(\Psi_\varepsilon) = \mathcal{L}(\theta) = 1$ establish that the neural networks $(\Psi_\varepsilon)_{\varepsilon,c\in(0,\infty)}$ satisfy (i). Furthermore note that Lemma 5.3, Lemma 5.4, and (b) ensure that for every $\varepsilon \in (0, \frac{B}{2m})$

$$\begin{aligned} \mathcal{M}(\Psi_\varepsilon) &= \mathcal{M}(\Pi_\varepsilon \circ \mathcal{P}(\Phi_\varepsilon^1, \Phi_\varepsilon^2, \ldots, \Phi_\varepsilon^m)) \\ &\leq 2\mathcal{M}(\Pi_\varepsilon) + \mathcal{M}(\mathcal{P}(\Phi_\varepsilon^1, \Phi_\varepsilon^2, \ldots, \Phi_\varepsilon^m)) + \mathcal{M}_{\mathcal{L}(\mathcal{P}(\Phi_\varepsilon^1, \Phi_\varepsilon^2, \ldots, \Phi_\varepsilon^m))}(\mathcal{P}(\Phi_\varepsilon^1, \Phi_\varepsilon^2, \ldots, \Phi_\varepsilon^m)) \\ &\leq 2C'm\left(|\ln(\varepsilon)| + m\ln(B) + \ln(m)\right) + \mathcal{M}(\Phi_\varepsilon^{\mathcal{P}}) + \mathcal{M}_{L_\varepsilon}(\Phi_\varepsilon^{\mathcal{P}}). \end{aligned} \tag{6.27}$$

This and the fact that for every $\varepsilon \in [\frac{B}{2m}, \infty)$ it holds that $\mathcal{M}(\Psi_\varepsilon) = \mathcal{M}(\theta) = 0$ imply the neural networks $(\Psi_\varepsilon)_{\varepsilon,c\in(0,\infty)}$ satisfy (ii). The proof of Proposition 6.4 is completed. □

Another way to use the multiplication results is to consider the approximation of smooth functions by polynomials. This can be done for functions of arbitrary dimension using the multivariate Taylor expansion (see [27] and [20, Thm. 2.3]). Such a direct approach, however, yields networks whose size depends exponentially on the dimension of the function. As our goal is to show that high dimensional functions with a tensor product structure can be approximated by networks with only polynomial dependence on the dimension, we only consider univariate smooth functions here. In the appendix we present a detailed and explicit construction of this Taylor approximation by neural networks.

Theorem 6.5. *Assume Setting 5.1, let $n \in \mathbb{N}$, $r \in (0, \infty)$, let $\varrho\colon \mathbb{R} \to \mathbb{R}$ be the ReLU activation function given by $\varrho(t) = \max\{0, t\}$, and let $B_1^n \subseteq C^n([0,1], \mathbb{R})$ be the set given by*

$$B_1^n = \left\{ f \in C^n([0,1], \mathbb{R})\colon \max_{k \in \{0,1,\ldots,n\}} \left[\sup_{t \in [0,1]} \left| f^{(k)}(t) \right| \right] \leq 1 \right\}. \tag{6.28}$$

Then there exist neural networks $(\Phi_{f,\varepsilon})_{f \in B_1^n, \varepsilon \in (0,\infty)} \subseteq \mathfrak{N}$ which satisfy

(i) $\displaystyle\sup_{f \in B_1^n, \varepsilon \in (0,\infty)} \left[\frac{\mathcal{L}(\Phi_{f,\varepsilon})}{\max\{r, |\ln(\varepsilon)|\}} \right] < \infty,$

(ii) $\displaystyle\sup_{f \in B_1^n, \varepsilon \in (0,\infty)} \left[\frac{\mathcal{M}(\Phi_{f,\varepsilon})}{\varepsilon^{-\frac{1}{n}} \max\{r, |\ln(\varepsilon)|\}} \right] < \infty,$ *and*

(iii) for every $f \in B_1^n$, $\varepsilon \in (0, \infty)$ that

$$\sup_{t \in [0,1]} |f(t) - [R_\varrho(\Phi_{f,\varepsilon})](t)| \leq \varepsilon. \tag{6.29}$$

For convenience of use we also provide the following more general corollary.

Corollary 6.6. *Assume Setting 5.1, let $r \in (0, \infty)$ and let $\varrho\colon \mathbb{R} \to \mathbb{R}$ be the ReLU activation function given by $\varrho(t) = \max\{0, t\}$. Let further the set $\mathcal{C}^n$ be given by $\mathcal{C}^n = \cup_{[a,b] \subseteq \mathbb{R}_+} C^n([a,b], \mathbb{R})$, and let $\|\cdot\|_{n,\infty} : \mathcal{C}^n \to [0, \infty)$ satisfy for every $[a,b] \subseteq \mathbb{R}_+$, $f \in C^n([a,b], \mathbb{R})$*

$$\|f\|_{n,\infty} = \max_{k \in \{0,1,\ldots,n\}} \left[\sup_{t \in [a,b]} \left| f^{(k)}(t) \right| \right]. \tag{6.30}$$

Then there exist neural networks $(\Phi_{f,\varepsilon})_{f \in \mathcal{C}^n, \varepsilon \in (0,\infty)} \subseteq \mathfrak{N}$ which satisfy

(i) $\displaystyle\sup_{f \in \mathcal{C}^n, \varepsilon \in (0,\infty)} \left[\frac{\mathcal{L}(\Phi_{f,\varepsilon})}{\max\{r, |\ln(\frac{\varepsilon}{\max\{1, b-a\} \|f\|_{n,\infty}})|\}} \right] < \infty,$

(ii) $\displaystyle\sup_{f \in \mathcal{C}^n, \varepsilon \in (0,\infty)} \left[\frac{\mathcal{M}(\Phi_{f,\varepsilon})}{\max\{1, b-a\} \|f\|_{n,\infty}^{\frac{1}{n}} \varepsilon^{-\frac{1}{n}} \max\{r, |\ln(\frac{\varepsilon}{\max\{1, b-a\} \|f\|_{n,\infty}})|\}} \right] < \infty,$ *and*

(iii) for every $[a,b] \subseteq \mathbb{R}_+$, $f \in C^n([a,b], \mathbb{R})$, $\varepsilon \in (0, \infty)$ that

$$\sup_{t \in [a,b]} |f(t) - [R_\varrho(\Phi_{f,\varepsilon})](t)| \leq \varepsilon. \tag{6.31}$$

7 DNN Expression Rates for High-Dimensional Basket prices

Now that we have established a number of general expression rate results, we can apply them to our specific problem. Using the regularity result (3.3) we obtain the following.

Corollary 7.1. *Assume Setting 5.1, let $n \in \mathbb{N}$, $r \in (0, \infty)$, $a \in (0, \infty)$, $b \in (a, \infty)$, let $\varrho\colon \mathbb{R} \to \mathbb{R}$ be the ReLU activation function given by $\varrho(t) = \max\{0, t\}$, let $f\colon (0, \infty) \to \mathbb{R}$ be as defined in (3.1),and let $h_{c,K}\colon [a,b] \to \mathbb{R}$, $c \in (0, \infty)$, $K \in [0, \infty)$, denote the functions which satisfy for every $c \in (0, \infty)$, $K \in [0, \infty)$, $x \in [a,b]$*

$$h_{c,K}(x) = f(\tfrac{K+c}{x}). \tag{7.1}$$

Then there exist neural networks $(\Phi_{\varepsilon,c,K})_{\varepsilon,c \in (0,\infty), K \in [0,\infty)} \subseteq \mathfrak{N}$ which satisfy *(i)* $\sup_{\varepsilon,c\in(0,\infty),K\in[0,\infty)} \left[\frac{\mathcal{L}(\Phi_{\varepsilon,c,K})}{\max\{r,|\ln(\varepsilon)|\}+\max\{0,\ln(K+c)\}}\right] < \infty,$

(ii) $\sup_{\varepsilon,c\in(0,\infty),K\in[0,\infty)} \left[\frac{\mathcal{M}(\Phi_{\varepsilon,c,K})}{(K+c+1)^{\frac{1}{n}}\varepsilon^{-\frac{1}{n^2}}}\right] < \infty,$ *and*

(iii) for every $\varepsilon, c \in (0,\infty)$, $K \in [0,\infty)$ *that*

$$\sup_{x\in[a,b]} |h_{c,K}(x) - [R_\varrho(\Phi_{\varepsilon,c,K})](x)| \leq \varepsilon. \tag{7.2}$$

Proof of Corollary 7.1. We observe Corollary 3.3 ensures the existence of a constant $C \in \mathbb{R}$ with

$$\max_{k\leq n} \sup_{x\in[a,b]} \left|h^{(k)}_{c,K}(x)\right| \leq C \max\{(K+c)^n, 1\}. \tag{7.3}$$

Moreover, observe for every $\varepsilon, c \in (0,\infty)$, $K \in [0,\infty)$ it holds

$$\begin{aligned} &\max\{r, |\ln(\tfrac{\varepsilon}{\max\{1,b-a\}C\max\{(K+c)^n,1\}})|\} \\ &\leq \max\{r,|\ln(\varepsilon)|\} + |\ln(\max\{1,b-a\})| + |\ln(C\max\{(K+c)^n,1\})| \\ &\leq \max\{r,|\ln(\varepsilon)|\} + \ln(\max\{1,b-a\}) + |\ln(C)| + |\ln(\max\{(K+c)^n,1\})| \\ &\leq \max\{r,|\ln(\varepsilon)|\} + \ln(\max\{1,b-a\}) + |\ln(C)| + n\max\{\ln(K+c),0\} \\ &\leq n(1+\max\{1,\tfrac{1}{r}\}(|\ln(C)|+\ln(\max\{1,b-a\})))(\max\{r,|\ln(\varepsilon)|\}+\max\{\ln(K+c),0\}). \end{aligned} \tag{7.4}$$

Furthermore, note for every $\varepsilon, c \in (0,\infty)$, $K \in [0,\infty)$ it holds

$$\begin{aligned} \left[\frac{\varepsilon}{\max\{1,b-a\}C\max\{(K+c)^n,1\}}\right]^{-\frac{1}{2n^2}} &= [\max\{1,b-a\}]^{-\frac{1}{2n^2}}\varepsilon^{-\frac{1}{2n^2}}C^{\frac{1}{2n^2}}\max\{(K+c)^{\frac{1}{2n}},1\} \\ &\leq [\max\{1,b-a\}]^{-\frac{1}{2n^2}}C^{\frac{1}{2n^2}}(K+c+1)^{\frac{1}{2n}}\varepsilon^{-\frac{1}{2n^2}}. \end{aligned} \tag{7.5}$$

Combining this, (7.3), (7.4) with Lemma A.2 and Corollary 6.6 (with $n \leftrightarrow 2n^2$ in the notation of Corollary 6.6) completes the proof of Corollary 7.1. □

We can then employ Proposition 6.4 in order to approximate the required tensor product.

Corollary 7.2. *Assume Setting 5.1, let* $\varrho\colon \mathbb{R} \to \mathbb{R}$ *be the ReLU activation function given by* $\varrho(t) = \max\{0,t\}$, *let* $n \in \mathbb{N}$, $a \in (0,\infty)$, $b \in (a,\infty)$, $(K_i)_{i\in\mathbb{N}} \subseteq [0, K_{\max})$, *and consider, for* $h_{c,K}\colon [a,b] \to \mathbb{R}$, $c \in (0,\infty)$, $K \in [0,K_{\max})$, *the functions which are, for every* $c \in (0,\infty)$, $K \in [0,K_{\max})$, $x \in [a,b]$, *given by*

$$h_{c,K}(x) = \tfrac{1}{\sqrt{2\pi}} \int_{-\infty}^{\ln(\frac{K+c}{x})} e^{-\frac{1}{2}r^2}\mathrm{d}r. \tag{7.6}$$

For any $c \in (0,\infty)$, $d \in \mathbb{N}$ *let the function* $F^d_c(x)\colon [a,b]^d \to \mathbb{R}$ *be given by*

$$F^d_c(x) = 1 - \left[\prod_{i=1}^d h_{c,K_i}(x_i)\right]. \tag{7.7}$$

Then there exist neural networks $(\Psi^d_{\varepsilon,c})_{\varepsilon,c\in(0,\infty),d\in\mathbb{N}} \subseteq \mathfrak{N}$ *which satisfy*

(i) $\sup_{\varepsilon,c\in(0,\infty),d\in\mathbb{N}} \left[\frac{\mathcal{L}(\Psi^d_{\varepsilon,c})}{\max\{1,\ln(d)\}(|\ln(\varepsilon)|+\ln(d)+1)+\ln(c+1)}\right] < \infty,$

(ii) $\sup_{\varepsilon,c\in(0,\infty),d\in\mathbb{N}} \left[\frac{\mathcal{M}(\Psi^d_{\varepsilon,c})}{(c+1)^{\frac{1}{n}}d^{1+\frac{1}{n}}\varepsilon^{-\frac{1}{n}}}\right] < \infty,$ *and*

(iii) for every $\varepsilon, c \in (0,\infty)$, $d \in \mathbb{N}$ that

$$\sup_{x\in[a,b]^d} \left|F_c^d(x) - \left[R_\varrho(\Psi_{\varepsilon,c}^d)\right](x)\right| \le \varepsilon. \tag{7.8}$$

Proof of Corollary 7.2. Throughout this proof assume Setting 5.2. Property Corollary 7.1 ensures there exist constants $b_L, b_M \in (0,\infty)$ and neural networks $\left(\Phi_{\eta,c}^i\right)_{\eta,c\in(0,\infty)} \subseteq \mathfrak{N}$, $i \in \mathbb{N}$ such that for every $i \in \mathbb{N}$ it holds

(a) $\sup_{\eta,c\in(0,\infty)} \left[\frac{\mathcal{L}(\Phi_{\eta,c}^i)}{\max\{1, |\ln(\eta)|\} + \max\{0, \ln(K_{\max} + c)\}}\right] < b_L$,

(b) $\sup_{\eta,c\in(0,\infty)} \left[\frac{\mathcal{M}(\Phi_{\eta,c}^i)}{(K_{\max} + c + 1)^{\frac{1}{n}} \eta^{-\frac{1}{n^2}}}\right] < b_M$, and

(c) for every $\eta, c \in (0,\infty)$ that

$$\sup_{x\in[a,b]} \left|h_{c,K_i}(x) - \left[R_\varrho(\Phi_{\eta,c}^i)\right](x)\right| \le \eta. \tag{7.9}$$

Furthermore, for every $c \in (0,\infty)$, $i \in \mathbb{N}$, $x \in [a,b]$ holds

$$|h_{c,K_i}(x)| = \left|\tfrac{1}{\sqrt{2\pi}} \int_{-\infty}^{\ln(\frac{K_i+c}{x})} e^{-\frac{1}{2}r^2}\mathrm{d}r\right| \le \tfrac{1}{\sqrt{2\pi}} \left|\int_{-\infty}^{\infty} e^{-\frac{1}{2}r^2}\mathrm{d}r\right| = 1. \tag{7.10}$$

Combining this with (a) and Proposition 6.4 and Lemma 5.4 implies there exist $C \in \mathbb{R}$ and neural networks $(\psi_{\eta,c}^d)_{\eta\in(0,\infty)} \subseteq \mathfrak{N}$, $c \in (0,\infty)$, $d \in \mathbb{N}$, such that for every $c \in (0,\infty)$, $d \in \mathbb{N}$ it holds

(A) $\mathcal{L}(\psi_{\eta,c}^d) \le C\ln(d)\,(|\ln(\eta)| + \ln(d)) + \max_{i\in\{1,2,\dots,d\}} \mathcal{L}(\Phi_{\eta,c}^i)$,

(B) $\mathcal{M}(\psi_{\eta,c}^d) \le Cd\,(|\ln(\eta)| + \ln(d)) + 4\sum_{i=1}^{d} \mathcal{M}(\Phi_{\eta,c}^i) + 8d \max_{i\in\{1,2,\dots,d\}} \mathcal{L}(\Phi_{\eta,c}^i)$, and

(C) for every $\eta \in (0,\infty)$ that

$$\sup_{x\in[a,b]^d} \left|\left[\prod_{i=1}^{d} h_{c,K_i}(x_i)\right] - \left[R_\varrho(\psi_{\eta,c}^d)\right](x)\right| \le 3d\eta. \tag{7.11}$$

Let $\lambda \in \mathcal{N}_1^{1,1}$ be the neural network given by $\lambda = ((-1,1))$, let $\theta \in \mathcal{N}_1^{1,1}$ be the neural network given by $\theta = (0,0)$, and let $(\Psi_{\varepsilon,c}^d)_{\varepsilon,c\in(0,\infty),d\in\mathbb{N}} \subseteq \mathfrak{N}$ be the neural networks given by

$$\Psi_{\varepsilon,c}^d = \begin{cases} \lambda \circ \psi_{\varepsilon/(3d),c}^d & : \varepsilon \le 2 \\ \theta & : \varepsilon > 2 \end{cases}. \tag{7.12}$$

Observe that this and (B) imply for every $\varepsilon \in (0,2]$, $c \in (0,\infty)$, $d \in \mathbb{N}$, $x \in [a,b]^d$ it holds

$$\begin{aligned} \left|F_c^d(x) - \left[R_\varrho(\Psi_{\varepsilon,c}^d)\right](x)\right| &= \left|\left(1 - \left[\prod_{i=1}^{d} h_{c,K_i}(x_i)\right]\right) - \left(1 - \left[R_\varrho(\psi_{\varepsilon/(3d),c}^d)\right](x)\right)\right| \\ &\le 3d\tfrac{\varepsilon}{3d} = \varepsilon. \end{aligned} \tag{7.13}$$

Moreover, (7.12) and (7.10) ensure for every $\varepsilon \in (2,\infty)$, $c \in (0,\infty)$, $d \in \mathbb{N}$, $x \in [a,b]^d$ it holds

$$\left|F_c^d(x) - \left[R_\varrho(\Psi_{\varepsilon,c}^d)\right](x)\right| = \left|\left(1 - \left[\prod_{i=1}^{d} h_{c,K_i}(x_i)\right]\right)\right| \tag{7.14}$$

This and (7.13) establish the neural networks $(\Psi^d_{\varepsilon,c})_{\varepsilon,c\in(0,\infty),d\in\mathbb{N}}$ satisfy (iii). Next observe that for every $c \in (0,\infty)$ it holds

$$\begin{aligned}\max\{0, \ln(K_{\max} + c)\} &\le \max\{0, \ln(\max\{1, K_{\max}\} + \max\{1, K_{\max}\}c)\} \\ &= \ln(\max\{1, K_{\max}\}(1 + c)) = \ln(\max\{1, K_{\max}\}) + \ln(1 + c) \\ &\le \ln(c + 1) + |\ln(K_{\max})|.\end{aligned} \tag{7.15}$$

Hence, we obtain that for every $\varepsilon, c \in (0,\infty)$, $d \in \mathbb{N}$ it holds

$$\begin{aligned}&\max\{1, |\ln(\tfrac{\varepsilon}{3d})|\} + \max\{0, \ln(K_{\max} + c)\} \\ &\le |\ln(\varepsilon)| + \ln(d) + \ln(3) + \ln(c + 1) + |\ln(K_{\max})| \\ &\le (\ln(3) + |\ln(K_{\max})|)\left[\max\{1, \ln(d)\}(|\ln(\varepsilon)| + \ln(d) + 1) + \ln(c + 1)\right].\end{aligned} \tag{7.16}$$

In addition, for every $\varepsilon, c \in (0,\infty)$, $d \in \mathbb{N}$ it holds

$$C \ln(d) \left(\left|\ln(\tfrac{\varepsilon}{3d})\right| + \ln(d)\right) \le 4C \left[\max\{1, \ln(d)\}(|\ln(\varepsilon)| + \ln(d) + 1) + \ln(c + 1)\right]. \tag{7.17}$$

Combining this with Lemma 5.3, (a), (A), and (7.16) yields

$$\begin{aligned}&\sup_{\substack{\varepsilon\in(0,2],c\in(0,\infty),\\ d\in\mathbb{N}}} \left[\frac{\mathcal{L}(\Psi^d_{\varepsilon,c})}{\max\{1, \ln(d)\}(|\ln(\varepsilon)| + \ln(d) + 1) + \ln(c + 1)}\right] \\ &\le \sup_{\substack{\varepsilon\in(0,2],c\in(0,\infty),\\ d\in\mathbb{N}}} \left[\frac{1 + C\ln(d)\left(\left|\ln(\tfrac{\varepsilon}{3d})\right| + \ln(d)\right) + \max_{i\in\{1,2,\dots,d\}} \mathcal{L}(\Phi^i_{\varepsilon/(3d),c})}{\max\{1, \ln(d)\}(|\ln(\varepsilon)| + \ln(d) + 1) + \ln(c + 1)}\right] \\ &\le 2 + 4C + (\ln(3) + |\ln(K_{\max})|)b_L < \infty.\end{aligned} \tag{7.18}$$

Moreover, (7.12) shows

$$\begin{aligned}&\sup_{\substack{\varepsilon\in(2,\infty),c\in(0,\infty),\\ d\in\mathbb{N}}} \left[\frac{\mathcal{L}(\Psi^d_{\varepsilon,c})}{\max\{1, \ln(d)\}(|\ln(\varepsilon)| + \ln(d) + 1) + \ln(c + 1)}\right] \\ &= \sup_{\substack{\varepsilon\in(2,\infty),c\in(0,\infty),\\ d\in\mathbb{N}}} \left[\frac{1}{\max\{1, \ln(d)\}(|\ln(\varepsilon)| + \ln(d) + 1) + \ln(c + 1)}\right] < \infty.\end{aligned} \tag{7.19}$$

This and (7.18) establish that $(\Psi^d_{\varepsilon,c})_{\varepsilon,c\in(0,\infty),d\in\mathbb{N}}$ satisfy (i). Next observe Lemma A.2 implies that

- for every $\varepsilon \in (0, 2]$ it holds

$$|\ln(\varepsilon)| \le \left[\sup_{\delta\in[\exp(-2n^2),2]} \ln(\delta)\right] \varepsilon^{-\frac{1}{n}} = 2n^2\varepsilon^{-\frac{1}{n}}, \tag{7.20}$$

- for every $d \in \mathbb{N}$ it holds

$$\ln(d) \le \left[\max_{k\in\{1,2,\dots,\exp(2n^2)\}} \ln(k)\right] d^{\frac{1}{n}} = 2n^2 d^{\frac{1}{n}}, \tag{7.21}$$

- and for every $c \in (0,\infty)$ it holds

$$\ln(c + 1) \le \left[\sup_{t\in(0,\exp(2n^2-1)]} \ln(t + 1)\right] (c + 1)^{\frac{1}{n}} = 2n^2(c + 1)^{\frac{1}{n}}. \tag{7.22}$$

For every $m \in \mathbb{N}$, $x_i \in [1,\infty)$, $i \in \{1, 2, \dots, m\}$, it holds

$$\sum_{i=1}^{m} x_i \le \prod_{i=1}^{m} (x_i + 1) \le 2^m \prod_{i=1}^{m} x_i. \tag{7.23}$$

Combining this with (7.20), (7.21), and (7.22) shows for every $\varepsilon \in (0,2]$, $d \in \mathbb{N}$, $c \in (0,\infty)$ it holds

$$\begin{aligned} 2Cd(|\ln(\tfrac{\varepsilon}{3d})| + \ln(d)) &\le 2Cd(|\ln(\varepsilon)| + 2\ln(d) + \ln(3) + \ln(c+1)) \\ &\le 4n^2Cd(2\varepsilon^{-\frac{1}{n}} + 2d^{\frac{1}{n}} + \ln(3) + (c+1)^{\frac{1}{n}}) \\ &\le 1024n^2C(c+1)^{\frac{1}{n}}d^{1+\frac{1}{n}}\varepsilon^{-\frac{1}{n}}. \end{aligned} \tag{7.24}$$

Furthermore, note (7.15), (7.20), (7.21), (7.22), and (7.23) ensure for every $\varepsilon \in (0,2]$, $d \in \mathbb{N}$, $c \in (0,\infty)$ it holds

$$\begin{aligned} &16d(\max\{1, |\ln(\tfrac{\varepsilon}{3d})|\} + \max\{0, \ln(K_{\max} + c)\}) \\ &\le 16d(|\ln(\varepsilon)| + \ln(d) + \ln(3) + \ln(c+1) + |\ln(K_{\max})|) \\ &\le 32n^2d(2\varepsilon^{-\frac{1}{n}} + d^{\frac{1}{n}} + (c+1)^{\frac{1}{n}} + \ln(3) + |\ln(K_{\max})|) \\ &\le 2048n^2(\ln(3) + |\ln(K_{\max})|)(c+1)^{\frac{1}{n}}d^{1+\frac{1}{n}}\varepsilon^{-\frac{1}{n}}. \end{aligned} \tag{7.25}$$

In addition, observe that for every $\varepsilon \in (0,2]$, $d \in \mathbb{N}$, $c \in (0,\infty)$ it holds

$$4d(K_{\max} + c + 1)^{\frac{1}{n}}(\tfrac{\varepsilon}{3d})^{-\frac{1}{n^2}} \le 96\max\{1, K_{\max}\}(c+1)^{\frac{1}{n}}d^{1+\frac{1}{n}}\varepsilon^{-\frac{1}{n}}. \tag{7.26}$$

Combining this with Lemma 5.3, (a), (b), (B), (7.24), and (7.25) yield

$$\begin{aligned} &\sup_{\substack{\varepsilon \in (0,2], c \in (0,\infty), \\ d \in \mathbb{N}}} \left[\frac{\mathcal{M}(\Psi^d_{\varepsilon,c})}{(c+1)^{\frac{1}{n}}d^{1+\frac{1}{n}}\varepsilon^{-\frac{1}{n}}} \right] \\ &\le \sup_{\substack{\varepsilon \in (0,2], c \in (0,\infty), \\ d \in \mathbb{N}}} \left[\frac{4 + 2Cd(|\ln(\frac{\varepsilon}{3d})| + \ln(d)) + 8\sum_{i=1}^{d} \mathcal{M}(\Phi^i_{\varepsilon/(3d),c}) + 16d \max_{i \in \{1,2,\ldots,d\}} \mathcal{L}(\Phi^i_{\varepsilon/(3d),c})}{(c+1)^{\frac{1}{n}}d^{1+\frac{1}{n}}\varepsilon^{-\frac{1}{n}}} \right] \\ &\le 8 + 1024n^2C + 96\max\{1, K_{\max}\}b_M + 2048n^2(\ln(3) + |\ln(K_{\max})|)b_L < \infty. \end{aligned} \tag{7.27}$$

Furthermore, note that (7.12) ensures

$$\sup_{\substack{\varepsilon \in (2,\infty), c \in (0,\infty), \\ d \in \mathbb{N}}} \left[\frac{\mathcal{M}(\Psi^d_{\varepsilon,c})}{(c+1)^{\frac{1}{n}}d^{1+\frac{1}{n}}\varepsilon^{-\frac{1}{n}}} \right] = \sup_{\substack{\varepsilon \in (2,\infty), c \in (0,\infty), \\ d \in \mathbb{N}}} \left[\frac{\mathcal{M}(\theta)}{(c+1)^{\frac{1}{n}}d^{1+\frac{1}{n}}\varepsilon^{-\frac{1}{n}}} \right] = 0. \tag{7.28}$$

This and (7.27) establish that the neural networks $(\Psi^d_{\varepsilon,c})_{\varepsilon, c \in (0,\infty), d \in \mathbb{N}}$ satisfy (ii). Thus the proof of Corollary 7.2 is completed. □

Finally, we add the quadrature estimates from Section 4 to achieve approximation with networks whose size only depends polynomially on the dimension of the problem.

Theorem 7.3. *Assume Setting 5.1, let $\varrho\colon \mathbb{R} \to \mathbb{R}$ be the ReLU activation function given by $\varrho(t) = \max\{0, t\}$, let $n \in \mathbb{N}$, $a \in (0,\infty)$, $b \in (a,\infty)$, $(K_i)_{i \in \mathbb{N}} \subseteq [0, K_{\max})$, and let $F_d\colon (0,\infty) \times [a,b]^d \to \mathbb{R}$, $d \in \mathbb{N}$, be the functions which satisfy for every $d \in \mathbb{N}$, $c \in (0,\infty)$, $x \in [a,b]^d$*

$$F_d(c,x) = 1 - \prod_{i=1}^{d} \left[\tfrac{1}{\sqrt{2\pi}} \int_{-\infty}^{\ln(\frac{K_i + c}{x_i})} e^{-\frac{1}{2}r^2} \mathrm{d}r \right]. \tag{7.29}$$

Then there exists neural networks $(\Gamma_{d,\varepsilon})_{\varepsilon \in (0,1], d \in \mathbb{N}} \in \mathfrak{N}$ which satisfy

(i) $\sup_{\varepsilon \in (0,1], d \in \mathbb{N}} \left[\frac{\mathcal{L}(\Gamma_{d,\varepsilon})}{\max\{1, \ln(d)\}\,(|\ln(\varepsilon)| + \ln(d) + 1)} \right] < \infty,$

(ii) $\sup_{\varepsilon\in(0,1],d\in\mathbb{N}}\left[\frac{\mathcal{M}(\Gamma_{d,\varepsilon})}{d^{2+\frac{1}{n}}\varepsilon^{-\frac{1}{n}}}\right]<\infty$, *and*

(iii) for every $\varepsilon\in(0,1]$, $d\in\mathbb{N}$ *that*

$$\sup_{x\in[a,b]^d}\left|\int_0^\infty F_d(c,x)\mathrm{d}c-[R_\varrho(\Gamma_{d,\varepsilon})](x)\right|\leq\varepsilon. \tag{7.30}$$

Proof of Theorem 7.3. Throughout this proof assume Setting 5.2, let $S_{b,n}\in\mathbb{R}$ be given by

$$S_{b,n}=2e^{2(4n+1)}(b+1)^{1+\frac{1}{4n}} \tag{7.31}$$

and let $N_{d,\varepsilon}\in\mathbb{R}$, $d\in\mathbb{N}$, $\varepsilon\in(0,1]$, be given by

$$N_{d,\varepsilon}=S_{b,n}d^{\frac{1}{4n}}\left[\tfrac{\varepsilon}{4}\right]^{-\frac{1}{4n}}. \tag{7.32}$$

Note Lemma 4.3 (with $4n\leftrightarrow n$, $F_x^d(c)\leftrightarrow F_d(x,c)$, $N_{d,\frac{\varepsilon}{2}}\leftrightarrow N_{d,\varepsilon}$, $Q_{d,\frac{\varepsilon}{2}}\leftrightarrow Q_{d,\varepsilon}$ in the notation of Lemma 4.3) ensures that there exist $Q_{d,\varepsilon}\in\mathbb{R}$, $c_{\varepsilon,j}^d\in(0,N_{d,\varepsilon})$, $w_{\varepsilon,j}^d\in[0,\infty)$, $j\in\{1,2,\ldots,Q_{d,\varepsilon}\}$, $d\in\mathbb{N}$, $\varepsilon\in(0,1]$ with

$$\sup_{\varepsilon\in(0,1],d\in\mathbb{N}}\left[\frac{Q_{d,\varepsilon}}{d^{1+\frac{1}{2n}}\varepsilon^{-\frac{1}{2n}}}\right]<\infty \tag{7.33}$$

and for every $d\in\mathbb{N}$, $\varepsilon\in(0,1]$ it holds

$$\sup_{x\in[a,b]^d}\left|\int_0^\infty F_d(c,x)\mathrm{d}c-\sum_{j=0}^{Q_{d,\varepsilon}}w_{\varepsilon,j}^dF_d(c_{\varepsilon,j}^d,x)\right|\leq\tfrac{\varepsilon}{2} \tag{7.34}$$

and

$$\sum_{j=1}^{Q_{d,\varepsilon}}w_{\varepsilon,j}^d=N_{d,\varepsilon}. \tag{7.35}$$

Furthermore, Corollary 7.2 (with $4n\leftrightarrow n$, $F^d_{c^d_{\varepsilon,j}}(x)\leftrightarrow F_d(x,c^d_{\varepsilon,j})$) ensures there exist neural networks $(\Psi^d_{\varepsilon,j})_{\varepsilon\in(0,\infty),d\in\mathbb{N},j\in\{1,2,\ldots,Q_{d,\varepsilon}\}}\subseteq\mathfrak{N}$ which satisfy

(a) $\sup_{\varepsilon\in(0,\infty),d\in\mathbb{N}}\left[\frac{\max_{j\in\{1,2,\ldots,Q_{d,\varepsilon}\}}\mathcal{L}(\Psi^d_{\varepsilon,j})}{\max\{1,\ln(d)\}\left(|\ln(\frac{\varepsilon}{2N_{d,\varepsilon}})|+\ln(d)+1\right)+\ln(N_{d,\varepsilon}+1)}\right]<\infty$,

(b) $\sup_{\varepsilon\in(0,\infty),d\in\mathbb{N}}\left[\frac{\max_{j\in\{1,2,\ldots,Q_{d,\varepsilon}\}}\mathcal{M}(\Psi^d_{\varepsilon,j})}{(N_{d,\varepsilon}+1)^{\frac{1}{4n}}d^{1+\frac{1}{4n}}\left[\frac{\varepsilon}{2N_{d,\varepsilon}}\right]^{-\frac{1}{4n}}}\right]<\infty$, and

(c) for every $\varepsilon\in(0,\infty)$, $d\in\mathbb{N}$ that

$$\sup_{x\in[a,b]^d}\left|F_d(c^d_{\varepsilon,j},x)-\left[R_\varrho(\Psi^d_{\varepsilon,j})\right](x)\right|\leq\tfrac{\varepsilon}{2N_{d,\varepsilon}}. \tag{7.36}$$

Let $\mathrm{Id}_{\mathbb{R}^d}\in\mathbb{R}^{d\times d}$, $d\in\mathbb{N}$, be the matrices given by $\mathrm{Id}_{\mathbb{R}^d}=\mathrm{diag}(1,1,\ldots,1)$, let $\nabla_{d,q}\in\mathcal{N}_1^{d,dq}$, $d,q\in\mathbb{N}$, be the neural networks given by

$$\nabla_{d,q}=\left(\left(\begin{pmatrix}\mathrm{Id}_d\\ \vdots\\ \mathrm{Id}_d\end{pmatrix},0\right)\right), \tag{7.37}$$

let $\Sigma_{d,\varepsilon} \in \mathcal{N}_1^{d,1}$, $d \in \mathbb{N}$, $\varepsilon \in (0,1]$, be the neural networks given by

$$\Sigma_{d,\varepsilon} = \left(\left(\begin{pmatrix} w^d_{\varepsilon,1} & w^d_{\varepsilon,2} & \dots & w^d_{\varepsilon,Q_{d,\varepsilon}} \end{pmatrix}, 0 \right) \right), \tag{7.38}$$

and let $(\Gamma_{d,\varepsilon})_{\varepsilon \in (0,1], d \in \mathbb{N}} \in \mathfrak{N}$ be the neural networks given by

$$\Gamma_{d,\varepsilon} = \Sigma_{d,\varepsilon} \circ \mathcal{P}(\Psi^d_{\varepsilon,1}, \Psi^d_{\varepsilon,2}, \dots, \Psi^d_{\varepsilon,Q_{d,\varepsilon}}) \circ \nabla_{d,Q_{d,\varepsilon}}. \tag{7.39}$$

Combining Lemma 5.3, Lemma 5.4, (7.34), (7.35), and (c) implies for every $\varepsilon \in (0,\infty)$ and $d \in \mathbb{N}$, $x \in [a,b]^d$ it holds

$$\begin{aligned} &\left| \int_0^\infty F_d(c,x) \mathrm{d}c - [R_\varrho(\Gamma_{d,\varepsilon})](x) \right| \\ &\le \left| \int_0^\infty F_d(c,x) \mathrm{d}c - \sum_{j=0}^{Q_{d,\varepsilon}} w^d_{\varepsilon,j} F_d(c^d_{\varepsilon,j}, x) \right| + \left| \sum_{j=0}^{Q_{d,\varepsilon}} w^d_{\varepsilon,j} F_d(c^d_{\varepsilon,j}, x) - [R_\varrho(\Gamma_{d,\varepsilon})](x) \right| \\ &\le \tfrac{\varepsilon}{2} + \left| \sum_{j=0}^{Q_{d,\varepsilon}} w^d_{\varepsilon,j} F_d(c^d_{\varepsilon,j}, x) - \sum_{j=0}^{Q_{d,\varepsilon}} w^d_{\varepsilon,j} \left[R_\varrho(\Psi^d_{\varepsilon,j}) \right](x) \right| \\ &\le \tfrac{\varepsilon}{2} + \sum_{j=0}^{Q_{d,\varepsilon}} w^d_{\varepsilon,j} \left| F_d(c^d_{\varepsilon,j}, x) - \left[R_\varrho(\Psi^d_{\varepsilon,j}) \right](x) \right| \le \tfrac{\varepsilon}{2} + N_{d,\varepsilon} \tfrac{\varepsilon}{2N_{d,\varepsilon}} = \varepsilon. \end{aligned} \tag{7.40}$$

This establishes that the neural networks $(\Gamma_{d,\varepsilon})_{\varepsilon \in (0,1], d \in \mathbb{N}}$ satisfy (iii). Next, observe for every $\varepsilon \in (0,\infty)$, $d \in \mathbb{N}$

$$\begin{aligned} &\max\{1, \ln(d)\} \left(|\ln(\frac{\varepsilon}{2N_{d,\varepsilon}})| + \ln(d) + 1 \right) + \ln(N_{d,\varepsilon} + 1) \\ &\le \max\{1, \ln(d)\} \left(|\ln(\varepsilon)| + \ln(d) + 3\ln(N_{d,\varepsilon}) + \ln(2) + 1 \right) \\ &\le \max\{1, \ln(d)\} \left(|\ln(\varepsilon)| + \ln(d) + 3 \left(\ln(S_{b,n}) + \frac{1}{4n} \ln(d) + \frac{1}{4n} |\ln(\varepsilon)| + \frac{1}{4n} \ln(4) \right) + 2 \right) \\ &\le \max\{1, \ln(d)\} \left(4|\ln(\varepsilon)| + 4\ln(d) + 3\ln(S_{b,n}) + 8 \right) \\ &\le (3\ln(S_{b,n}) + 8) \max\{1, \ln(d)\} \left(|\ln(\varepsilon)| + \ln(d) + 1 \right). \end{aligned} \tag{7.41}$$

Combining this with Lemma 5.3, Lemma 5.4, and (a) implies

$$\begin{aligned} &\sup_{\varepsilon \in (0,1], d \in \mathbb{N}} \left[\frac{\mathcal{L}(\Gamma_{d,\varepsilon})}{\max\{1, \ln(d)\} \left(|\ln(\varepsilon)| + \ln(d) + 1 \right)} \right] \\ &\le \sup_{\varepsilon \in (0,1], d \in \mathbb{N}} \left[\frac{\mathcal{L}(\Sigma_{d,\varepsilon}) + \max_{j \in \{1,2,\dots,Q_{d,\varepsilon}\}} \mathcal{L}(\Psi^d_{\varepsilon,j}) + \mathcal{L}(\nabla_{d,Q_{d,\varepsilon}})}{\max\{1, \ln(d)\} \left(|\ln(\varepsilon)| + \ln(d) + 1 \right)} \right] \\ &\le 2 + \sup_{\varepsilon \in (0,1], d \in \mathbb{N}} \left[\frac{\max_{j \in \{1,2,\dots,Q_{d,\varepsilon}\}} \mathcal{L}(\Psi^d_{\varepsilon,j})}{\max\{1, \ln(d)\} \left(|\ln(\varepsilon)| + \ln(d) + 1 \right)} \right] \\ &\le 2 + (3\ln(S_{b,n}) + 8) \sup_{\varepsilon \in (0,\infty), d \in \mathbb{N}} \left[\frac{\max_{j \in \{1,2,\dots,Q_{d,\varepsilon}\}} \mathcal{L}(\Psi^d_{\varepsilon,j})}{\max\{1, \ln(d)\} \left(|\ln(\frac{\varepsilon}{2N_{d,\varepsilon}})| + \ln(d) + 1 \right) + \ln(N_{d,\varepsilon} + 1)} \right] \\ &< \infty. \end{aligned} \tag{7.42}$$

This establishes $(\Gamma_{d,\varepsilon})_{\varepsilon\in(0,1],d\in\mathbb{N}}$ satisfy (i). In addition, for every $\varepsilon \in (0,\infty)$, $d \in \mathbb{N}$ it holds

$$\begin{aligned}(N_{d,\varepsilon}+1)^{\frac{1}{4n}} d^{1+\frac{1}{4n}} \left[\frac{\varepsilon}{2N_{d,\varepsilon}}\right]^{-\frac{1}{4n}} &\leq 4N_{d,\varepsilon}^{\frac{1}{2n}} d^{1+\frac{1}{4n}} \varepsilon^{-\frac{1}{4n}} \\ &\leq 4\left[S_{b,n} d^{\frac{1}{4n}} \left[\tfrac{\varepsilon}{4}\right]^{-\frac{1}{4n}}\right]^{\frac{1}{2n}} d^{1+\frac{1}{4n}} \varepsilon^{-\frac{1}{4n}} \\ &\leq 16 S_{b,n} d^{1+\frac{1}{4n}+\frac{1}{4n^2}} \varepsilon^{-(\frac{1}{4n}+\frac{1}{8n^2})} \\ &\leq 16 S_{b,n} d^{1+\frac{1}{2n}} \varepsilon^{-\frac{1}{2n}}.\end{aligned} \tag{7.43}$$

Combining this with Lemma 5.3, Lemma 5.4, (7.33), (b), and the fact that for every $\psi \in \mathfrak{N}$ which satisfies $\min_{l\in\{1,2,\ldots,\mathcal{L}(\psi)\}} \mathcal{M}_l(\psi) > 0$ it holds $\mathcal{L}(\psi) \leq \mathcal{M}(\psi)$ ensures

$$\begin{aligned}&\sup_{\varepsilon\in(0,1],d\in\mathbb{N}} \left[\frac{\mathcal{M}(\Gamma_{d,\varepsilon})}{d^{(2+\frac{1}{n})}\varepsilon^{-\frac{1}{n}}}\right] \\ &\leq \sup_{\varepsilon\in(0,1],d\in\mathbb{N}} \left[\frac{2\mathcal{M}(\Sigma_{d,\varepsilon}) + 4\left(2\sum_{j=1}^{Q_{d,\varepsilon}} \mathcal{M}(\Psi^d_{\varepsilon,j}) + 4Q_{d,\varepsilon} \max_{j\in\{1,2,\ldots,Q_{d,\varepsilon}\}} \mathcal{L}(\Psi^d_{\varepsilon,j})\right) + 4\mathcal{M}(\nabla_{d,Q_{d,\varepsilon}})}{d^{(2+\frac{1}{n})}\varepsilon^{-\frac{1}{n}}}\right] \\ &\leq \sup_{\varepsilon\in(0,1],d\in\mathbb{N}} \left[\frac{24Q_{d,\varepsilon}\max_{j\in\{1,2,\ldots,Q_{d,\varepsilon}\}} \mathcal{M}(\Psi^d_{\varepsilon,j})}{d^{(2+\frac{1}{n})}\varepsilon^{-\frac{1}{n}}}\right] + \sup_{\varepsilon\in(0,1],d\in\mathbb{N}} \left[\frac{2Q_{d,\varepsilon}+4dQ_{d,\varepsilon}}{d^{(2+\frac{1}{n})}\varepsilon^{-\frac{1}{n}}}\right] \\ &\leq 24\left(\sup_{\varepsilon\in(0,1],d\in\mathbb{N}} \left[\frac{Q_{d,\varepsilon}}{d^{(1+\frac{1}{2n})}\varepsilon^{-\frac{1}{2n}}}\right]\right)\left(\sup_{\varepsilon\in(0,1],d\in\mathbb{N}} \left[\frac{\max_{j\in\{1,2,\ldots,Q_{d,\varepsilon}\}} \mathcal{M}(\Psi^d_{\varepsilon,j})}{d^{(1+\frac{1}{2n})}\varepsilon^{-\frac{1}{2n}}}\right]\right) \\ &\quad + 4\sup_{\varepsilon\in(0,1],d\in\mathbb{N}} \left[\frac{Q_{d,\varepsilon}}{d^{(1+\frac{1}{n})}\varepsilon^{-\frac{1}{n}}}\right] \\ &\leq 24\left(\sup_{\varepsilon\in(0,1],d\in\mathbb{N}} \left[\frac{Q_{d,\varepsilon}}{d^{(1+\frac{1}{2n})}\varepsilon^{-\frac{1}{2n}}}\right]\right)\left(1 + 16S_{b,n}\sup_{\varepsilon\in(0,1],d\in\mathbb{N}} \left[\frac{\max_{j\in\{1,2,\ldots,Q_{d,\varepsilon}\}} \mathcal{M}(\Psi^d_{\varepsilon,j})}{(N_{d,\varepsilon}+1)^{\frac{1}{4n}} d^{1+\frac{1}{4n}} \left[\frac{\varepsilon}{2N_{d,\varepsilon}}\right]^{-\frac{1}{4n}}}\right]\right) \\ &< \infty.\end{aligned} \tag{7.44}$$

This establishes the neural networks $(\Gamma_{d,\varepsilon})_{\varepsilon\in(0,1],d\in\mathbb{N}}$ satisfy (ii). The proof of Theorem 7.3 is thus completed. □

References

[1] Barron, A. R. Universal approximation bounds for superpositions of a sigmoidal function. *IEEE Transactions on Information theory 39*, 3 (1993), 930–945.

[2] Beck, C., Becker, S., Grohs, P., Jaafari, N., and Jentzen, A. Solving stochastic differential equations and Kolmogorov equations by means of deep learning. *arXiv:1806.00421* (2018).

[3] Beck, C., E, W., and Jentzen, A. Machine learning approximation algorithms for high-dimensional fully nonlinear partial differential equations and second-order backward stochastic differential equations. *arXiv:1709.05963* (2017), 56 pages.

[4] Bölcskei, H., Grohs, P., Kutyniok, G., and Petersen, P. Optimal approximation with sparsely connected deep neural networks. *arXiv preprint arXiv:1705.01714* (2017).

[5] Chiani, M., Dardari, D., and Simon, M. K. New exponential bounds and approximations for the computation of error probability in fading channels. *IEEE Trans. Wireless Communications 2*, 4 (2003), 840–845.

[6] Chui, C., Li, X., and Mhaskar, H. Neural networks for localized approximation. *Mathematics of Computation 63*, 208 (1994), 607–623.

[7] Cybenko, G. Approximation by superpositions of a sigmoidal function. *Mathematics of control, signals and systems 2*, 4 (1989), 303–314.

[8] E, W., Han, J., and Jentzen, A. Deep learning-based numerical methods for high-dimensional parabolic partial differential equations and backward stochastic differential equations. *Commun. Math. Stat. 5*, 4 (2017), 349–380.

[9] E, W., and Yu, B. The deep Ritz method: a deep learning-based numerical algorithm for solving variational problems. *Commun. Math. Stat. 6*, 1 (2018), 1–12.

[10] Freidlin, M. *Functional integration and partial differential equations*, vol. 109 of *Annals of Mathematics Studies*. Princeton University Press, Princeton, NJ, 1985.

[11] Fujii, M., Takahashi, A., and Takahashi, M. Asymptotic Expansion as Prior Knowledge in Deep Learning Method for high dimensional BSDEs. *arXiv:1710.07030* (2017), 16 pages.

[12] Goodfellow, I., Bengio, Y., Courville, A., and Bengio, Y. *Deep learning*, vol. 1. MIT press Cambridge, 2016.

[13] Hairer, M., Hutzenthaler, M., and Jentzen, A. Loss of regularity for Kolmogorov equations. *Annals of Probability 2015, Vol. 43, No. 2, 468-527* (Mar. 2015).

[14] Han, J., Jentzen, A., and E, W. Overcoming the curse of dimensionality: Solving high-dimensional partial differential equations using deep learning. *arXiv:1707.02568* (2017), 13 pages.

[15] Henry-Labordere, P. Deep Primal-Dual Algorithm for BSDEs: Applications of Machine Learning to CVA and IM. 16 pages. Available at SSRN: https://ssrn.com/abstract=3071506.

[16] Hornik, K., Stinchcombe, M., and White, H. Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks. *Neural networks 3*, 5 (1990), 551–560.

[17] Khoo, Y., Lu, J., and Ying, L. Solving parametric PDE problems with artificial neural networks. *arXiv:1707.03351* (2017), 12 pages.

[18] Kwok, Y.-K. *Mathematical models of financial derivatives*, second ed. Springer Finance. Springer, Berlin, 2008.

[19] Levy, D. Introduction to Numerical Analysis, 2010.

[20] Mhaskar, H. N. Neural Networks for optimal approximation of smooth and analytic functions. *Neural Computation 8* (1996), 164–177.

[21] Perekreshtenko, D., Grohs, P., Elbrächter, D., and Bölcskei, H. The Universal Approximation Power of Finite Width Deep ReLU Networks. submitted.

[22] Petersen, P., and Voigtlaender, F. Optimal approximation of piecewise smooth functions using deep ReLU neural networks, Sept. 2017.

[23] Pinkus, A. Approximation theory of the MLP model in neural networks. *Acta Numer. 8* (1999), 143–195.

[24] Schwab, C., and Zech, J. Deep Learning in High Dimension: neural network expression rates for generalized polynomial chaos expressions in UQ. *Research Report 2017* (2017). (to appear in Analysis and Applications (Singap.)(2018)).

[25] Sirignano, J., and Spiliopoulos, K. DGM: A deep learning algorithm for solving partial differential equations. *arXiv:1708.07469* (2017), 16 pages.

[26] Telgarsky, M. Representation Benefits of Deep Feedforward Networks. *CoRR abs/1509.08101* (2015).

[27] Yarotsky, D. Error bounds for approximations with deep ReLU networks. *Neural Networks 94* (2017), 103–114.

A Additional Proofs

A.1 Complementary Distribution Formula

Lemma A.1 (Complementary distribution function formula)**.** *Let $\mu\colon \mathcal{B}([0,\infty)) \to [0,\infty]$ be a sigma-finite measure. Then*

$$\int_0^\infty x\,\mu(dx) = \int_0^\infty \mu([x,\infty))\,dx = \int_0^\infty \mu((x,\infty))\,dx. \tag{A.1}$$

Proof of Lemma 2.2. First, observe that

$$\begin{aligned}\int_0^\infty x\,\mu(dx) &= \int_0^\infty \left[\int_0^x dy\right]\mu(dx) = \int_0^\infty \left[\int_0^\infty \mathbb{1}_{(-\infty,x]}(y)\,dy\right]\mu(dx)\\ &= \int_0^\infty \int_0^\infty \mathbb{1}_{[y,\infty)}(x)\,dy\,\mu(dx).\end{aligned} \tag{A.2}$$

Next observe that the fact that function

$$[0,\infty)^2 \ni (x,y) \mapsto \mathbb{1}_{[y,\infty)}(x) \in \mathbb{R} \tag{A.3}$$

is $(\mathcal{B}([0,\infty)) \otimes \mathcal{B}([0,\infty)))/\mathcal{B}(\mathbb{R})$-measurable and the hypothesis that μ is a sigma-finite measure allows us to apply Fubini's theorem to obtain

$$\int_0^\infty \int_0^\infty \mathbb{1}_{[y,\infty)}(x)\,dy\,\mu(dx) = \int_0^\infty \int_0^\infty \mathbb{1}_{[y,\infty)}(x)\,\mu(dx)\,dy = \int_0^\infty \mu([y,\infty))\,dy\,. \tag{A.4}$$

Combining this with (A.2) demonstrates for every $\varepsilon \in (0,\infty)$

$$\begin{aligned}\int_0^\infty x\,\mu(dx) &= \int_0^\infty \mu([y,\infty))\,dy \geq \int_0^\infty \mu((y,\infty))\,dy\\ &\geq \int_0^\infty \mu([y+\varepsilon,\infty))\,dy = \int_\varepsilon^\infty \mu([y,\infty))\,dy.\end{aligned} \tag{A.5}$$

Beppo Levi's monotone convergence theorem hence establishes

$$\begin{aligned}\int_0^\infty x\,\mu(dx) &= \int_0^\infty \mu([y,\infty))\,dy \geq \int_0^\infty \mu((y,\infty))\,dy\\ &\geq \sup_{\varepsilon\in(0,\infty)} \left[\int_\varepsilon^\infty \mu([y,\infty))\,dy\right]\\ &= \sup_{\varepsilon\in(0,\infty)} \left[\int_0^\infty \mu([y,\infty))\,\mathbb{1}_{(\varepsilon,\infty)}(y)\,dy\right] = \int_0^\infty \mu([y,\infty))\,dy.\end{aligned} \tag{A.6}$$

The proof of Lemma A.1 is thus completed. ∎

A.2 Technical Lemma

Lemma A.2. *It holds for every* $r \in (0,\infty)$, $t \in (0, \exp(-2r^2)]$ *that*

$$|\ln(t)| \le t^{-1/r} \tag{A.7}$$

and for every $r \in (0,\infty)$, $t \in [\exp(2r^2), \infty)$ *that*

$$\ln(t) \le t^{1/r}. \tag{A.8}$$

Proof of Lemma A.2. First, observe that for every $r \in (0,\infty)$, $y \in [2r^2, \infty)$ it holds that

$$\exp\Big(\frac{y}{r}\Big) = \sum_{k=0}^{\infty}\left[\frac{y^k}{k!r^k}\right] \ge \frac{y^2}{2!r^2} = y\left[\frac{y}{2r^2}\right] \ge y. \tag{A.9}$$

This implies that for every $r \in (0,\infty)$, $x \in [\exp(2r^2), \infty)$ it holds that

$$x^{1/r} = \exp\Big(\ln\Big(x^{1/r}\Big)\Big) = \exp\Big(\tfrac{\ln(x)}{r}\Big) \ge \ln(x). \tag{A.10}$$

Hence, we obtain that for every $r \in (0,\infty)$, $t \in (0, \exp(-2r^2)] \subseteq (0,1]$ it holds that

$$t^{-1/r} = \left[\tfrac{1}{t}\right]^{1/r} \ge \ln(\tfrac{1}{t}) = |\ln(t)|\,. \tag{A.11}$$

This completes the proof of Lemma A.2. ☐

A.3 Proof of Lemma 6.1

Proof of Lemma 6.1. The proof follows [27]. We provide it in order to provide values of constants in the bounds on depth and width, and to reveal the dependence on the scaling parameter B. Throughout this proof let $\theta \in \mathcal{N}_1^{1,1}$ be the neural network given by $\theta = (0,0)$, let $g_s \colon [0,1] \to [0,1]$, $s \in \mathbb{N}$, be the functions which satisfy for every $s \in \mathbb{N}$, $t \in [0,1]$ that

$$g_s(t) = \begin{cases} 2t & : s = 1, t < \frac{1}{2} \\ 2 - 2t & : s = 1, t \ge \frac{1}{2} \\ g_1(g_{s-1}(t)) & : s \ge 1 \end{cases}, \tag{A.12}$$

and let $f_m \colon [0,1] \to [0,1]$, $m \in \mathbb{N}$, be the functions which satisfy for every $m \in \mathbb{N}$, $k \in \{0,1,\dots,2^m\}$, $x \in \left[\frac{k}{2^m}, \frac{k+1}{2^m}\right]$ that

$$f_m(x) = \left[\frac{2k+1}{2^m}\right]x - \frac{k^2+k}{2^{2m}}. \tag{A.13}$$

We claim for every $s \in \mathbb{N}$, $k \in \{0,1,\dots,2^{s-1}-1\}$ it holds

$$g_s(x) = \begin{cases} 2^s(x - \frac{2k}{2^s}) & : x \in \left[\frac{2k}{2^s}, \frac{2k+1}{2^s}\right] \\ 2^s(\frac{2k+2}{2^s} - x) & : x \in \left[\frac{2k+1}{2^s}, \frac{2k+2}{2^s}\right] \end{cases}. \tag{A.14}$$

We now prove (A.14) by induction on $s \in \mathbb{N}$. Equation (A.12) establishes (A.14) in the base case $s = 1$. For the induction step $\mathbb{N} \ni s \to s+1 \in \{2,3,\dots\}$ observe that (A.12) implies for every $s \in \mathbb{N}$, $l \in \{0,1,\dots,2^{s-1}-1\}$ that

(a) it holds for every $x \in \left[\frac{2l}{2^s}, \frac{2l+(1/2)}{2^s}\right]$

$$\begin{aligned} g_{s+1}(x) &= g(g_s(x)) = g(2^s(x - \tfrac{2l}{2^s})) = 2\left[2^s(x - \tfrac{2l}{2^s})\right] \\ &= 2^{s+1}(x - \tfrac{2l}{2^s}) = 2^{s+1}(x - \tfrac{2(2l)}{2^{s+1}}). \end{aligned} \tag{A.15}$$

(b) it holds for every $x \in \left[\frac{2l+(1/2)}{2^s}, \frac{2l+1}{2^s}\right]$

$$\begin{aligned} g_{s+1}(x) &= g(g_s(x)) = g(2^s(x - \tfrac{2l}{2^s})) = 2 - 2\left[2^s(x - \tfrac{2l}{2^s})\right] \\ &= 2 - 2^{s+1}x + 4l = 2^{s+1}(\tfrac{4l+2}{2^{s+1}} - x) \\ &= 2^{s+1}(\tfrac{2(2l+1)}{2^{s+1}} - x). \end{aligned} \tag{A.16}$$

(c) it holds for every $x \in \left[\frac{2l+1}{2^s}, \frac{2l+(3/2)}{2^s}\right]$

$$\begin{aligned} g_{s+1}(x) &= g(g_s(x)) = g(2^s(\tfrac{2l+1}{2^s} - x)) = 2 - 2\left[2^s(\tfrac{2l+1}{2^s} - x)\right] \\ &= 2 - 2(2l+1) + 2^{s+1}x = 2^{s+1}x - 2(2l) \\ &= 2^{s+1}(x - \tfrac{2(2l)}{2^{s+1}}). \end{aligned} \tag{A.17}$$

(d) it holds for every $x \in \left[\frac{2l+(3/2)}{2^s}, \frac{2l+2}{2^s}\right]$

$$\begin{aligned} g_{s+1}(x) &= g(g_s(x)) = g(2^s(\tfrac{2l+1}{2^s} - x)) = 2\left[2^s(\tfrac{2l+1}{2^s} - x)\right] \\ &= 2^{s+1}(\tfrac{2l+1}{2^s} - x) = 2^{s+1}(\tfrac{2(2l+1)}{2^{s+1}} - x). \end{aligned} \tag{A.18}$$

Next observe that for every $s \in \mathbb{N}$, $k \in \{0, 1, \ldots, 2^s - 1\}$ there exists $l \in \{0, 1, \ldots, 2^{s-1} - 1\}$ such that

$$\left[\tfrac{2k}{2^{s+1}}, \tfrac{2k+1}{2^{s+1}}\right] = \left[\tfrac{2l}{2^s}, \tfrac{2l+(1/2)}{2^s}\right] \quad \text{or} \quad \left[\tfrac{2k}{2^{s+1}}, \tfrac{2k+1}{2^{s+1}}\right] = \left[\tfrac{2l+1}{2^s}, \tfrac{2l+(3/2)}{2^s}\right]. \tag{A.19}$$

Furthermore, for every $s \in \mathbb{N}$, $k \in \{0, 1, \ldots, 2^s - 1\}$ there exists $l \in \{0, 1, \ldots, 2^{s-1} - 1\}$ such that

$$\left[\tfrac{2k+1}{2^{s+1}}, \tfrac{2k+2}{2^{s+1}}\right] = \left[\tfrac{2l+(1/2)}{2^s}, \tfrac{2l+1}{2^s}\right] \quad \text{or} \quad \left[\tfrac{2k+1}{2^{s+1}}, \tfrac{2k+2}{2^{s+1}}\right] = \left[\tfrac{2l+(3/2)}{2^s}, \tfrac{2l+2}{2^s}\right]. \tag{A.20}$$

Combining this with (A.15), (A.16), (A.17), (A.18), and (A.19) completes the induction step $\mathbb{N} \ni s \to s+1 \in \{2, 3, \ldots\}$ and thus establishes the claim (A.14).

Next, for every $m \in \mathbb{N}$, $k \in \{0, 1, \ldots, 2^{m-1}\}$ it holds

$$f_{m-1}(\tfrac{2k}{2^m}) - f_m(\tfrac{2k}{2^m}) = f_{m-1}(\tfrac{k}{2^{m-1}}) - f_m(\tfrac{2k}{2^m}) = \left[\tfrac{k}{2^{m-1}}\right]^2 - \left[\tfrac{2k}{2^m}\right]^2 = 0. \tag{A.21}$$

In addition, note that (A.13) implies that for every $m \in \mathbb{N}$, $k \in \{0, 1, \ldots, 2^m - 1\}$ it holds

$$\begin{aligned} f_{m-1}(\tfrac{2k+1}{2^m}) &= f_{m-1}\left(\tfrac{k+\frac{1}{2}}{2^{m-1}}\right) = \left[\frac{2k+1}{2^{m-1}}\right]\frac{k+\frac{1}{2}}{2^{m-1}} - \frac{k^2+k}{2^{2(m-1)}} \\ &= \frac{(2k+1)(k+\frac{1}{2}) - (k^2+k)}{2^{2m-2}} = \frac{k^2+k+\frac{1}{2}}{2^{2m-2}} = \frac{4k^2+4k+2}{2^{2m}} \end{aligned} \tag{A.22}$$

and

$$f_m(\tfrac{2k+1}{2^m}) = \left[\frac{2(2k+1)+1}{2^m}\right]\frac{2k+1}{2^m} - \frac{(2k+1)^2+(2k+1)}{2^{2m}} = \frac{4k^2+4k+1}{2^{2m}}. \tag{A.23}$$

For every $m \in \mathbb{N}$, $k \in \{0, 1, \ldots, 2^m - 1\}$ it holds

$$f_{m-1}(\tfrac{2k+1}{2^m}) - f_m(\tfrac{2k+1}{2^m}) = \frac{4k^2+4k+2}{2^{2m}} - \frac{4k^2+4k+1}{2^{2m}} = \frac{1}{2^{2m}}. \tag{A.24}$$

Combining this with (A.14), (A.13), and (A.21) demonstrates that for every $m \in \mathbb{N}$, $x \in [0, 1]$ it holds

$$f_{m-1}(x) - f_m(x) = 2^{-2m} g_m(x). \tag{A.25}$$

The fact that for every $x \in [0,1]$ it holds that $f_0(x) = x$ therefore implies that for every $m \in \mathbb{N}_0$, $x \in [0,1]$ it holds

$$f_m(x) = x - \sum_{s=1}^{m} 2^{-2s} g_s(x). \tag{A.26}$$

We observe f_m is the affine, linear interpolant of the twice continuously differentiable function $[0,1] \ni x \mapsto x^2 \in [0,1]$ at the points $\frac{k}{2^m}$, $k \in \{0, 1, \ldots, 2^m\}$. This establishes that for every $m \in \mathbb{N}$

$$\begin{aligned}
\sup_{x\in[0,1]} \left|x^2 - f_m(x)\right| &= \max_{k\in\{0,1,\ldots,2^m\}} \left(\sup_{x\in\left[\frac{k}{2^m},\frac{k+1}{2^m}\right]} \left|x^2 - f_m(x)\right| \right) \\
&\le \max_{k\in\{0,1,\ldots,2^m\}} \left(\frac{\left[\frac{k+1}{2^m} - \frac{k}{2^m}\right]^2}{8} \max_{x\in\left[\frac{k}{2^m},\frac{k+1}{2^m}\right]} \left| \frac{\mathrm{d}^2}{\mathrm{d}t^2} \left[x^2\right] \right| \right) \\
&\le \max_{k\in\{0,1,\ldots,2^m\}} \left(\frac{1}{8} \left[\frac{1}{2^m}\right]^2 \max_{x\in\left[\frac{k}{2^m},\frac{k+1}{2^m}\right]} |2| \right) \\
&= 2^{-2m-2}.
\end{aligned} \tag{A.27}$$

Let $(A_k, b_k) \in \mathbb{R}^{4\times4} \times \mathbb{R}^4$, $k \in \mathbb{N}$, be the matrix-vector tuples which satisfy for every $k \in \mathbb{N}$

$$A_k = \begin{pmatrix} 2 & -4 & 2 & 0 \\ 2 & -4 & 2 & 0 \\ 2 & -4 & 2 & 0 \\ -2^{-2k+3} & 2^{-2k+4} & -2^{-2k+3} & 1 \end{pmatrix} \quad \text{and} \quad b_k = \begin{pmatrix} 0 \\ -\frac{1}{2} \\ -1 \\ 0 \end{pmatrix}, \tag{A.28}$$

let $\varphi_m \in \mathfrak{N}$, $m \in \mathbb{N}$, be the neural networks which satisfy $\varphi_1 = (1, 0)$ and, for every $m \in \mathbb{N}$,

$$\varphi_m = \left(\left(\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ -\frac{1}{2} \\ -1 \\ 0 \end{pmatrix} \right), (A_2, b_2), \ldots, (A_{m-1}, b_{m-1}), \left(\begin{pmatrix} -2^{-2m+3} \\ 2^{-2m+4} \\ -2^{-2m+3} \\ 1 \end{pmatrix}^T, 0 \right) \right). \tag{A.29}$$

Let further $r^k \colon \mathbb{R} \to \mathbb{R}$, $k \in \mathbb{N}$ denote the function which satisfies for every $x \in \mathbb{R}$

$$(r_1^1(x), r_2^1(x), r_3^1(x), r_4^1(x)) = r^1(x) = \varrho^*(x, x - \tfrac{1}{2}, x - 1, x) \tag{A.30}$$

and for every $x \in \mathbb{R}$, $k \in \mathbb{N}$

$$(r_1^k(x), r_2^k(x), r_3^k(x), r_4^k(x)) = r^k(x) = \varrho^*(A_k r_{k-1}(x) + b_k). \tag{A.31}$$

We claim that for every $k \in \{1, 2, \ldots, m-1\}$, $x \in [0,1]$ it holds

(a)

$$2r_1^k(x) - 4r_2^k(x) + 2r_3^k(x) = g_k(x) \tag{A.32}$$

and

(b)

$$r_4^k(x) = x - \sum_{j=1}^{k-1} 2^{-2j} g_j(x). \tag{A.33}$$

We prove (a) and (b) by induction over $k \in \{1, 2, \dots, m-1\}$. For the base case $k = 1$ we note that for every $x \in [0, 1]$ it holds

$$g_1(x) = 2\varrho(x) - 4\varrho(x - \tfrac{1}{2}) + 2\varrho(x - 1). \tag{A.34}$$

Hence, we obtain that for every $x \in [0, 1]$ it holds

$$2r_1^1(x) - 4r_2^1(x) + 2r_3^1(x) = 2\varrho(x) - 4\varrho(x - \tfrac{1}{2}) + 2\varrho(x - 1) = g_1(x). \tag{A.35}$$

Furthermore, note that for every $x \in [0, 1]$ it holds that $r_4^1(x) = x$. This and (A.35) establish the base case $k = 1$. For the induction step $\{1, 2, \dots, m-2\} \ni k-1 \to k \in \{2, 3, \dots, m-1\}$ observe that (A.34) ensures for every $x \in [0, 1]$, $k \in \{2, 3, \dots, m-1\}$, with $g_{k-1}(x) = 2r_1^{k-1}(x) - 4r_2^{k-1}(x) + 2r_3^{k-1}(x)$, it holds

$$\begin{aligned} 2r_1^k(x) - 4r_2^k(x) + 2r_3^k(x) = {} & \ 2\varrho(2r_1^{k-1}(x) - 4r_2^{k-1}(x) + 2r_3^{k-1}(x)) \\ & -4\varrho(2r_1^{k-1}(x) - 4r_2^{k-1}(x) + 2r_3^{k-1}(x) - \tfrac{1}{2}) \\ & +2\varrho(2r_1^{k-1}(x) - 4r_2^{k-1}(x) + 2r_3^{k-1}(x) - 1) \\ = {} & \ g_1(2r_1^{k-1}(x) - 4r_2^{k-1}(x) + 2r_3^{k-1}(x)) \\ = {} & \ g_1(g_{k-1}(k)) = g_k(x). \end{aligned} \tag{A.36}$$

Induction thus establishes (a). Moreover note that (A.13) and (A.26) for every $k \in \mathbb{N}$, $x \in [0, 1]$ it holds

$$x - \sum_{j=1}^{k-1} 2^{-2j} g_j(x) = f_{k-1}(x) \geq 0. \tag{A.37}$$

Combining this with (A.34) implies that for every $x \in [0, 1]$, $k \in \{2, 3, \dots, m-1\}$ with $g_{k-1}(x) = 2r_1^{k-1}(x) - 4r_2^{k-1}(x) + 2r_3^{k-1}(x)$ and $r_4^{k-1}(x) = x - \sum_{j=1}^{k-2} 2^{-2j} g_j(x)$ it holds

$$\begin{aligned} r_4^k(x) &= \varrho(-2^{-2k+3} r_1^{k-1}(x) + 2^{-2k+4} r_2^{k-1}(x) - 2^{-2k+3} r_3^{k-1}(x) + r_4^{k-1}(x)) \\ &= \varrho(x - \sum_{j=1}^{k-2} 2^{-2j} g_j(x) - g_{k-1}(x)) = \varrho(x - \sum_{j=1}^{k-1} 2^{-2j} g_j(x)) \\ &= x - \sum_{j=1}^{k-1} 2^{-2j} g_j(x). \end{aligned} \tag{A.38}$$

Induction thus establishes (b). Next observe that (a) and (b) that for every $m \in \mathbb{N}$, $x \in [0, 1]$ it holds

$$\begin{aligned} [R_\varrho(\varphi_m)](x) &= -2^{-2m+3} r_1^{m-1}(x) + 2^{-2m+4} r_2^{m-1}(x) - 2^{-2m+3} r_3^{m-1}(x) + r_4^{m-1}(x) \\ &= -2^{-2(m-1)} \left(2r_1^{m-1}(x) - 4r_2^{m-1}(x) + 2r_3^{m-1}(x)\right) + x - \sum_{j=1}^{m-2} 2^{-2j} g_j(x) \\ &= x - \left[\sum_{j=1}^{m-2} 2^{-2j} g_j(x)\right] - 2^{-2(m-1)} g_{m-1}(x) = x - \sum_{j=1}^{m-1} 2^{-2j} g_j(x). \end{aligned} \tag{A.39}$$

Combining this with (A.26) establishes that for every $m \in \mathbb{N}$, $x \in [0, 1]$ it holds

$$[R_\varrho(\varphi_m)](x) = f_{m-1}(x). \tag{A.40}$$

This and (A.27) imply that for every $m \in \mathbb{N}$ it holds

$$\sup_{x \in [0,1]} \left| x^2 - [R_\varrho(\varphi_m)](x) \right| \leq 2^{-2m}. \tag{A.41}$$

Furthermore, observe that by construction it holds for every $m \in \mathbb{N}$

$$\mathcal{L}(\varphi_m) = m \quad \text{and} \quad \mathcal{M}(\varphi_m) = \max\{1, 10 + 15(m-2)\} \leq 15m. \tag{A.42}$$

Let $(\sigma_\varepsilon)_{\varepsilon \in (0,\infty)} \subseteq \mathfrak{N}$ be the neural networks which satisfy for $\varepsilon \in (0,1)$

$$\sigma_\varepsilon = \varphi_{\lceil \frac{1}{2} |\log_2(\varepsilon)| \rceil} \tag{A.43}$$

and for every $\varepsilon \in [1,\infty)$ that $\sigma_\varepsilon = \theta$. Observe that for every $\varepsilon \in [1,\infty)$ it holds

$$\sup_{x \in [0,1]} \left| x^2 - [R_\varrho(\sigma_\varepsilon)](x) \right| = \sup_{x \in [0,1]} \left| x^2 - [R_\varrho(\theta)](x) \right| \leq 1 \leq \varepsilon. \tag{A.44}$$

In addition note for every $\varepsilon \in (0,1)$ it holds

$$\begin{aligned} \sup_{x \in [0,1]} \left| x^2 - [R_\varrho(\sigma_\varepsilon)](x) \right| &= \sup_{x \in [0,1]} \left| x^2 - [R_\varrho(\varphi_{\lceil \frac{1}{2} |\log_2(\varepsilon)| \rceil})](x) \right| \\ &\leq 2^{-2\lceil \frac{1}{2} |\log_2(\varepsilon)| \rceil} \leq 2^{-2(\frac{1}{2}|\log_2(\varepsilon)|)} = 2^{\log_2(\varepsilon)} = \varepsilon. \end{aligned} \tag{A.45}$$

Moreover, observe that (A.42) implies for every $\varepsilon \in (0,1)$ it holds

$$\mathcal{L}(\sigma_\varepsilon) = \mathcal{L}(\varphi_{\lceil \frac{1}{2} |\log_2(\varepsilon)| \rceil}) = \left\lceil \tfrac{1}{2} |\log_2(\varepsilon)| \right\rceil \tag{A.46}$$

and

$$\mathcal{M}(\sigma_\varepsilon) = \mathcal{M}(\varphi_{\lceil \frac{1}{2} |\log_2(\varepsilon)| \rceil}) \leq 15 \left\lceil \tfrac{1}{2} |\log_2(\varepsilon)| \right\rceil. \tag{A.47}$$

Furthermore, for every $\varepsilon \in [1,\infty)$ it holds $\mathcal{L}(\sigma_\varepsilon) = \mathcal{L}(\theta) = 1$ and $\mathcal{M}(\sigma_\varepsilon) = \mathcal{M}(\theta) = 0$. This completes the proof of Lemma 6.1. ☐

A.4 Proof of Lemma 6.2

Proof of Lemma 6.2. Throughout this proof assume Setting 5.2, let $\theta \in \mathcal{N}_1^{1,1}$ be the neural network given by $\theta = (0,0)$, let $\alpha_1, \alpha_2, \alpha_{1,2} \in \mathcal{N}_2^{2,2,1}$ be the neural networks given by

$$\begin{aligned} \alpha_1 &= \left((\begin{pmatrix} 1 & 0 \\ -1 & 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \end{pmatrix}), (\begin{pmatrix} \frac{1}{2B} & \frac{1}{2B} \end{pmatrix}, 0) \right), \\ \alpha_2 &= \left((\begin{pmatrix} 0 & 1 \\ 0 & -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \end{pmatrix}), (\begin{pmatrix} \frac{1}{2B} & \frac{1}{2B} \end{pmatrix}, 0) \right), \\ \alpha_{1,2} &= \left((\begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \end{pmatrix}), (\begin{pmatrix} \frac{1}{2B} & \frac{1}{2B} \end{pmatrix}, 0) \right), \end{aligned} \tag{A.48}$$

and let $\Sigma \in \mathcal{N}_1^{3,1}$ be the neural network given by $\Sigma = \left((\begin{pmatrix} 2B^2 & -2B^2 & -2B^2 \end{pmatrix}, 0) \right)$. Observe that Lemma 6.1 ensures the existence of neural networks $(\sigma_\varepsilon)_{\varepsilon \in (0,\infty)} \subseteq \mathfrak{N}$ which satisfy Lemma 6.1, (i) – (iv). Let $(\mu_\varepsilon)_{\varepsilon \in (0,\infty)} \subseteq \mathfrak{N}$ be the neural networks which satisfy for every $\varepsilon \in (0,\infty)$

$$\mu_\varepsilon = \begin{cases} \Sigma \circ \mathcal{P}\left(\sigma_{\varepsilon/6B^2} \circ \alpha_{1,2}, \sigma_{\varepsilon/6B^2} \circ \alpha_1, \sigma_{\varepsilon/6B^2} \circ \alpha_2 \right) & : \varepsilon < B^2 \\ \theta & : \varepsilon \geq B^2 \end{cases}. \tag{A.49}$$

Note first that for every $\varepsilon \in [B^2,\infty)$ it holds

$$\sup_{x,y \in [-B,B]} |xy - [R_\varrho(\mu_\varepsilon)](x,y)| = \sup_{x,y \in [-B,B]} |xy - [R_\varrho(\theta)](x,y)| = \sup_{x,y \in [-B,B]} |xy - 0| = B^2 \leq \varepsilon. \tag{A.50}$$

Next observe that for every $(x,y) \in \mathbb{R}^2$ it holds

$$\begin{aligned} [R_\varrho(\alpha_1)](x,y) &= \tfrac{1}{2B} \varrho(x) + \tfrac{1}{2B} \varrho(-x) = \tfrac{1}{2B} |x|, \; [R_\varrho(\alpha_2)](x,y) = \tfrac{1}{2B} |y|, \\ [R_\varrho(\alpha_{1,2})](x,y) &= \tfrac{1}{2B} |x+y|. \end{aligned} \tag{A.51}$$

Furthermore, for every $(x, y, z) \in \mathbb{R}^3$ holds $[R_\varrho(\Sigma)](x, y, z) = 2B^2x - 2B^2y - 2B^2z$. Combining this with Lemma 5.3, Lemma 5.4, (A.49), and (A.51) establishes that for every $\varepsilon \in (0, B^2)$, $(x, y) \in [-B, B]^2$ it holds

$$[R_\varrho(\mu_\varepsilon)](x, y) = 2B^2 \left([R_\varrho(\sigma_{\varepsilon/6B^2})] \left(\tfrac{|x+y|}{2B}\right) - [R_\varrho(\sigma_{\varepsilon/6B^2})] \left(\tfrac{|x|}{2B}\right) - [R_\varrho(\sigma_{\varepsilon/6B^2})] \left(\tfrac{|y|}{2B}\right) \right). \tag{A.52}$$

With Lemma 6.1, Item iv, (A.52) establishes (v). In addition note that Lemma 6.1 demonstrates for every $\varepsilon \in (0, \infty)$ it holds

$$\begin{aligned} &\sup_{z \in [-2B, 2B]} \left| \tfrac{1}{2} z^2 - 2B^2 \left[[R_\varrho(\sigma_{\varepsilon/6B^2})] \left(\tfrac{|z|}{2B}\right) \right] \right| \\ &= \sup_{z \in [-2B, 2B]} \left| 2B^2 \left[\tfrac{|z|}{2B}\right]^2 - 2B^2 \left[[R_\varrho(\sigma_{\varepsilon/6B^2})] \left(\tfrac{|z|}{2B}\right) \right] \right| \\ &= 2B^2 \left[\sup_{t \in [0,1]} \left| t^2 - \left[[R_\varrho(\sigma_{\varepsilon/6B^2})] (t) \right] \right| \right] \le 2B^2 \left[\frac{\varepsilon}{6B^2} \right] = \frac{\varepsilon}{3}. \end{aligned} \tag{A.53}$$

This and (A.52) establish that for every $\varepsilon \in (0, B^2)$ it holds

$$\begin{aligned} &\sup_{x, y \in [-B, B]} |xy - [R_\varrho(\mu_\varepsilon)](x, y)| \\ &= \sup_{x, y \in [-B, B]} \left| \frac{1}{2} \left[(x+y)^2 - x^2 - y^2 \right] - [R_\varrho(\mu_\varepsilon)](x, y) \right| \\ &\le \tfrac{\varepsilon}{3} + \tfrac{\varepsilon}{3} + \tfrac{\varepsilon}{3} = \varepsilon. \end{aligned} \tag{A.54}$$

Next observe that $\mathcal{L}(\alpha_1) = \mathcal{L}(\alpha_2) = \mathcal{L}(\alpha_{1,2}) = 2$ and $\mathcal{L}(\Sigma) = 1$. Combining this with Lemma 5.3, Lemma 5.4, and Lemma 6.1(i) ensures for every $\varepsilon \in (0, B^2)$

$$\begin{aligned} \mathcal{L}(\mu_\varepsilon) &= \mathcal{L}(\Sigma) + \max\{\mathcal{L}(\sigma_{\varepsilon/6B^2}) + \mathcal{L}(\alpha_{1,2}), \mathcal{L}(\sigma_{\varepsilon/6B^2}) + \mathcal{L}(\alpha_1), \mathcal{L}(\sigma_{\varepsilon/6B^2}) + \mathcal{L}(\alpha_2)\} \\ &\le \tfrac{1}{2} \left| \log_2(\tfrac{\varepsilon}{6B^2}) \right| + 4 = \tfrac{1}{2} \log_2(\tfrac{6B^2}{\varepsilon}) + 4 \\ &= \tfrac{1}{2} (\log_2(\tfrac{1}{\varepsilon}) + 2\log_2(B) + 3) + 4 \\ &= \tfrac{1}{2} \log_2(\tfrac{1}{\varepsilon}) + \log_2(B) + 6. \end{aligned} \tag{A.55}$$

Combining $\mathcal{M}(\alpha_1) = \mathcal{M}(\alpha_2) = 4$ and $\mathcal{M}(\alpha_{1,2}) = 6$ with Lemma 5.3, Lemma 6.1(ii), and (A.48) demonstrate for every $\varepsilon \in (0, B^2)$ it holds

$$\begin{aligned} \mathcal{M}(\sigma_{\varepsilon/6B^2} \circ \alpha_{1,2}) &\le 2(\mathcal{M}(\sigma_{\varepsilon/6B^2}) + \mathcal{M}(\alpha_{1,2})) \\ &\le 2(15(\tfrac{1}{2} \left| \log_2(\tfrac{\varepsilon}{6B^2}) \right| + 1) + 6) \\ &\le 15(\log_2(\tfrac{1}{\varepsilon}) + 2\log_2(B) + 3) + 42 \\ &= 15\log_2(\tfrac{1}{\varepsilon}) + 30\log_2(B) + 87 \end{aligned} \tag{A.56}$$

and analogously $\mathcal{M}(\sigma_{\varepsilon/6B^2} \circ \alpha_1) = \mathcal{M}(\sigma_{\varepsilon/6B^2} \circ \alpha_2) \le 15\log_2(\frac{1}{\varepsilon}) + 30\log_2(B) + 83$. This, and $\mathcal{M}(\Sigma) = 3$, Lemma 5.3 and Lemma 5.4 imply that for every $\varepsilon \in (0, B^2)$ it holds

$$\begin{aligned} \mathcal{M}(\mu_\varepsilon) &= 2\left(\mathcal{M}(\Sigma) + \left[\mathcal{M}(\sigma_{\varepsilon/6B^2} \circ \alpha_{1,2}) + \mathcal{M}(\sigma_{\varepsilon/6B^2} \circ \alpha_1) + \mathcal{M}(\sigma_{\varepsilon/6B^2} \circ \alpha_2)\right]\right) \\ &\le 90\log_2(\tfrac{1}{\varepsilon}) + 180\log_2(B) + 467. \end{aligned} \tag{A.57}$$

Moreover, for every $\varepsilon \in (B^2, \infty)$ it holds $\mathcal{L}(\mu_\varepsilon) = 1$ and $\mathcal{M}(\mu_\varepsilon) = 0$. Next, observe Lemma 5.3 and Lemma 5.4 demonstrate that for every $\varepsilon \in (0, \infty)$ it holds that

$$\mathcal{M}_1(\mu_\varepsilon) = \mathcal{M}_1(\mathcal{P}\left(\sigma_{\varepsilon/6B^2} \circ \alpha_{1,2}, \sigma_{\varepsilon/6B^2} \circ \alpha_1, \sigma_{\varepsilon/6B^2} \circ \alpha_2\right)) \le \mathcal{M}_1(\alpha_{1,2}) + \mathcal{M}_1(\alpha_1) + \mathcal{M}_1(\alpha_2) = 14 \tag{A.58}$$

and

$$\mathcal{M}_{\mathcal{L}(\mu_\varepsilon)}(\mu_\varepsilon) = \mathcal{M}(\Sigma) = 3. \tag{A.59}$$

This completes the proof of Lemma 6.2. □

A.5 Proof of Theorem 6.5

Proof of Theorem 6.5. Throughout this proof assume Setting 5.2, let $h_{N,j}\colon \mathbb{R} \to \mathbb{R}$, $N \in \mathbb{N}$, $j \in \{0,1,\dots,N\}$, be the functions which satisfy for every $N \in \mathbb{N}$, $j \in \{0,1,\dots,N\}$, $x \in \mathbb{R}$

$$h_{N,j}(x) = \begin{cases} Nx + 1 - j & : \frac{j-1}{N} \leq x \leq \frac{j}{N} \\ -Nx + 1 + j & : \frac{j}{N} \leq x \leq \frac{j+1}{N} \\ 0 & : \text{else} \end{cases}, \tag{A.60}$$

let $T_{f,N,j}\colon \mathbb{R} \to \mathbb{R}$, $f \in B_1^n$, $N \in \mathbb{N}$, $j \in \{0,1,\dots,N\}$, be the functions which satisfy for every $f \in B_1^n$, $N \in \mathbb{N}$, $j \in \{0,1,\dots,N\}$, $x \in [0,1]$

$$T_{f,N,j}(x) = \sum_{k=0}^{n-1} \frac{f^{(k)}(\frac{j}{N})}{k!}(x - \tfrac{j}{N})^k. \tag{A.61}$$

For every $f \in B_1^n$, let $f_N : \mathbb{R} \to \mathbb{R}$, $N \in \mathbb{N}$ denote functions which satisfy for every $N \in \mathbb{N}$, $x \in [0,1]$

$$f_N(x) = \sum_{j=0}^{N} h_{N,j}(x) T_{f,N,j}(x). \tag{A.62}$$

Observe that Taylor's theorem (with Lagrange remainder term) ensures that for every $f \in B_1^n$, $N \in \mathbb{N}$, $j \in \{0,1,\dots,N\}$, $x \in [\max\{0, \frac{j-1}{N}\}, \min\{1, \frac{j+1}{N}\}]$

$$\begin{aligned} |f(x) - T_{f,N,j}(x)| &\leq \tfrac{1}{n!} \left| x - \tfrac{j}{N} \right|^n \sup_{\xi \in [\max\{0,\frac{j-1}{N}\},\min\{1,\frac{j+1}{N}\}]} \left| f^{(n)}(\xi) \right| \\ &\leq \tfrac{1}{n!} N^{-n} \max_{k \in \{0,1,\dots,n\}} \left[\sup_{t \in [0,1]} \left| f^{(k)}(t) \right| \right] \leq \tfrac{1}{n!} N^{-n}. \end{aligned} \tag{A.63}$$

Moreover, for every $N \in \mathbb{N}$, $x \in [0,1]$, $j \notin \{\lceil Nx \rceil - 1, \lceil Nx \rceil\}$ it holds that $h_{N,j}(x) = 0$. We obtain for every $N \in \mathbb{N}$ and $x \in [0,1]$

$$\sum_{j=0}^{N} h_{N,j}(x) T_{f,N,j}(x) = h_{N,\lceil Nx \rceil - 1}(x) T_{f,N,\lceil Nx \rceil - 1}(x) + h_{N,\lceil Nx \rceil}(x) T_{f,N,\lceil Nx \rceil}(x). \tag{A.64}$$

Furthermore, (A.60) implies for every $N \in \mathbb{N}$, $j \in \{1,\dots,N-1\}$, $x \in [\frac{j-1}{N}, \frac{j}{N}]$ holds

$$h_{N,j-1}(x) + h_{N,j}(x) = -Nx + 1 + (j-1) + Nx + 1 - j = 1. \tag{A.65}$$

Combining this with (A.62), (A.63), and (A.64) establishes that for every $f \in B_1^n$, $N \in \mathbb{N}$, $x \in [0,1]$

$$\begin{aligned} &|f(x) - f_N(x)| \\ &= \left| f(x) - \sum_{j=0}^{N} h_{N,j}(x) T_{f,N,j}(x) \right| \\ &= \left| f(x) - \left(h_{N,\lceil Nx \rceil - 1}(x) T_{f,N,\lceil Nx \rceil - 1}(x) + h_{N,\lceil Nx \rceil}(x) T_{f,N,\lceil Nx \rceil}(x) \right) \right| \\ &\leq \left| h_{N,\lceil Nx \rceil - 1}(x) f(x) - h_{N,\lceil Nx \rceil - 1}(x) T_{f,N,\lceil Nx \rceil - 1}(x) \right| \\ &\quad + \left| h_{N,\lceil Nx \rceil}(x) f(x) - h_{N,\lceil Nx \rceil}(x) T_{f,N,\lceil Nx \rceil}(x) \right| \\ &= h_{N,\lceil Nx \rceil - 1}(x) \left| f(x) - T_{f,N,\lceil Nx \rceil - 1}(x) \right| + h_{N,\lceil Nx \rceil}(x) \left| f(x) - T_{f,N,\lceil Nx \rceil}(x) \right| \\ &\leq h_{N,\lceil Nx \rceil - 1}(x) \left[\tfrac{1}{n!} N^{-n} \right] + h_{N,\lceil Nx \rceil}(x) \left[\tfrac{1}{n!} N^{-n} \right] = \tfrac{1}{n!} N^{-n}. \end{aligned} \tag{A.66}$$

We now realize this local Taylor approximation using neural networks. To this end, note that Theorem 6.3 ensures that there exist $C \in \mathbb{R}$ and neural networks $(\Pi_\eta^k)_{\eta \in (0,\infty)}$, $k \in \mathbb{N} \cap [2,\infty)$ which satisfy

(A) $\mathcal{L}(\Pi_\eta^k) \leq C \ln(k) \left(|\ln(\eta)| + k\ln(3) + \ln(k)\right)$,

(B) $\mathcal{M}(\Pi_\eta^k) \leq Ck \left(|\ln(\eta)| + k\ln(3) + \ln(k)\right)$,

(C) $\sup_{x\in[-3,3]^k} \left| \left[\prod_{i=1}^k x_i\right] - \left[R_\varrho(\Pi_\eta^k)\right](x) \right| \leq \eta$ and

(D) $R_\varrho\left[\Pi_\eta^k\right](x_1, x_2, \ldots, x_k) = 0$, if there exists $i \in \{1, 2, \ldots, k\}$ with $x_i = 0$.

To complete the proof, we introduce the following neural networks:

- $\nabla_{N,j,k} \in \mathcal{N}_1^{k,1}$, $N \in \mathbb{N}$, $j \in \{0, 1, \ldots, N\}$, $k \in \{2, 3, \ldots, n-1\}$ given by

$$\nabla_{N,j,k} = \left(\left(\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}, \begin{pmatrix} -\frac{j}{N} \\ \vdots \\ -\frac{j}{N} \end{pmatrix} \right) \right), \tag{A.67}$$

- $\xi_{\varepsilon,N,j}^k \in \mathfrak{N}$, $\varepsilon \in (0,\infty)$, $N \in \mathbb{N}$, $j \in \{0, 1, \ldots, N\}$, $k \in \{1, 2, \ldots, n-1\}$, given by

$$\xi_{\varepsilon,N,j}^k = \begin{cases} (1,0) & : k = 1 \\ \Pi_{\varepsilon/8e}^k \circ \nabla_{N,j,k} & : k > 1 \end{cases}, \tag{A.68}$$

- $\Sigma_{f,N,j} \in \mathcal{N}_1^{1,n-1}$, $f \in B_1^n$, $N \in \mathbb{N}$, $j \in \{0, 1, \ldots, N\}$ given by

$$\Sigma_{f,N,j} = \left(\left(\begin{pmatrix} \frac{f^{(n-1)}(\frac{j}{N})}{(n-1)!} & \frac{f^{(n-2)}(\frac{j}{N})}{(n-2)!} & \ldots & \frac{f^{(1)}(\frac{j}{N})}{(1)!} \end{pmatrix}, f(\tfrac{j}{N}) \right) \right), \tag{A.69}$$

- $\tau_{f,\varepsilon,N,j} \in \mathfrak{N}$, $f \in B_1^n$, $\varepsilon \in (0,\infty)$, $N \in \mathbb{N}$, $j \in \{0, 1, \ldots, N\}$ given by

$$\tau_{f,\varepsilon,N,j} = \Sigma_{f,N,j} \circ \mathcal{P}(\xi_{\varepsilon,N,j}^{n-1}, \xi_{\varepsilon,N,j}^{n-2}, \ldots, \xi_{\varepsilon,N,j}^1) \circ \nabla_{1,0,n-1}, \tag{A.70}$$

- $\chi_{N,j} \in \mathcal{N}_2^{1,3,1}$, $N \in \mathbb{N}$, $j \in \{0, 1, \ldots, N\}$ given by

$$\chi_{N,j} = \left(\left(\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} -(j-1)/N \\ -j/N \\ -(j+1)/N \end{pmatrix} \right), \left(\begin{pmatrix} 1 & -2 & 1 \end{pmatrix}, 0 \right) \right) \tag{A.71}$$

- $\lambda_N \in \mathcal{N}_1^{1,N+1}$, $N \in \mathbb{N}$ given by

$$\lambda_N = \left(\left(\begin{pmatrix} 1 & \ldots & 1 \end{pmatrix}, 0 \right) \right), \tag{A.72}$$

- $\psi_{f,\varepsilon,N,j} \in \mathfrak{N}$, $f \in B_1^n$, $\varepsilon \in (0,\infty)$, $N \in \mathbb{N}$, $j \in \{0, 1, \ldots, N\}$ given by

$$\psi_{f,\varepsilon,N,j} = \Pi_{\varepsilon/8}^2 \circ \mathcal{P}(\chi_{N,j}, \tau_{f,\varepsilon,N,j}), \tag{A.73}$$

- $\varphi_{f,\varepsilon,N} \in \mathfrak{N}$, $f \in B_1^n$, $N \in \mathbb{N}$, $\varepsilon \in (0,\infty)$ given by

$$\varphi_{f,\varepsilon,N} = \lambda_N \circ \mathcal{P}\left(\psi_{f,\varepsilon,N,1}, \psi_{f,\varepsilon,N,2}, \ldots, \psi_{f,\varepsilon,N,N}\right) \circ \nabla_{1,0,2N+2}. \tag{A.74}$$

With these networks, we note Lemma 5.3, Lemma 5.4, (C), (A.67) and (A.68) ensure that for every $N \in \mathbb{N}$, $\varepsilon \in (0,\infty)$, $j \in \{0,1,\dots,N\}$, $k \in \{2,3,\dots,n-1\}$

$$\begin{aligned}
&\sup_{x\in[0,1]} \left|(x-\tfrac{j}{N})^k - \left[R_\varrho(\xi^k_{\varepsilon,N,j})\right](x)\right| \\
&\le \sup_{x\in[0,1]} \left|(x-\tfrac{j}{N})^k - \left[R_\varrho(\Pi^k_{\varepsilon/8e})\right]\left(\left[R_\varrho(\nabla_{N,j,k})\right](x)\right)\right| \\
&\le \sup_{x\in[0,1]} \left|\left[\prod_{i=1}^k (x-\tfrac{j}{N})^k\right] - \left[R_\varrho(\Pi^k_{\varepsilon/8e})\right](x-\tfrac{j}{N}, x-\tfrac{j}{N},\dots,x-\tfrac{j}{N})\right| \\
&\le \sup_{x\in[-1,1]^k} \left|\left[\prod_{i=1}^k x_i\right] - \left[R_\varrho(\Pi^k_{\varepsilon/8e})\right](x)\right| \le \tfrac{\varepsilon}{8e}
\end{aligned} \tag{A.75}$$

and

$$\sup_{x\in[0,1]} \left|(x-\tfrac{j}{N}) - \left[R_\varrho(\xi^1_{\varepsilon,N,j})\right](x)\right| = 0. \tag{A.76}$$

Moreover, Lemma 5.3, Lemma 5.4, (A.67), (A.68), (A.69), and (A.70) demonstrate that for every $f \in B^n_1$, $N \in \mathbb{N}$, $\varepsilon \in (0,\infty)$, $j \in \{0,1,\dots,N\}$, $x \in [0,1]$ it holds

$$\left[R_\varrho(\tau_{f,\varepsilon,N,j})\right](x) = \sum_{k=1}^{n-1}\left[\frac{f^{(k)}(\frac{j}{N})}{k!}\left[R_\varrho(\xi^k_{\varepsilon,N,j})\right](x)\right] + f(\tfrac{j}{N}). \tag{A.77}$$

Combining this with (A.61), (A.70), (A.75) and (A.75) establishes that for every $f \in B^n_1$, $N \in \mathbb{N}$, $\varepsilon \in (0,\infty)$, $j \in \{0,1,\dots,N\}$, $x \in [0,1]$ it holds

$$\begin{aligned}
&\left|T_{f,N,j}(x) - \left[R_\varrho(\tau_{f,\varepsilon,N,j})\right](x)\right| \\
&= \left|\left(\sum_{k=0}^{n-1}\frac{f^{(k)}(\frac{j}{N})}{k!}(x-\tfrac{j}{N})^k\right) - \left(\sum_{k=1}^{n-1}\left[\frac{f^{(k)}(\frac{j}{N})}{k!}\left[R_\varrho(\xi^k_{\varepsilon,N,j})\right](x)\right] + f(\tfrac{j}{N})\right)\right| \\
&\le \sum_{k=1}^{n-1}\left(\frac{f^{(k)}(\frac{j}{N})}{k!}\left|(x-\tfrac{j}{N})^k - \left[R_\varrho(\xi^k_{\varepsilon,N,j})\right](x)\right|\right) \\
&\le \frac{\varepsilon}{8e}\sum_{k=1}^{n-1}\frac{f^{(k)}(\frac{j}{N})}{k!} \le \frac{\varepsilon}{8e}\left(\sum_{k=1}^{\infty}\frac{1}{k!}\right) \le \frac{\varepsilon}{8}.
\end{aligned} \tag{A.78}$$

Next, (A.71) ensures for every $N \in \mathbb{N}$, $j \in \{0,1,\dots,N\}$, $x \in [0,1]$

$$[R_\varrho(\chi_{N,j})](x) = \varrho(x-\tfrac{j-1}{N}) - 2\varrho(x-\tfrac{j}{N}) + \varrho(x-\tfrac{j+1}{N}) = h_{N,j}(x). \tag{A.79}$$

Now (A.78) and Taylor's Theorem imply for every $f \in B^n_1$, $N \in \mathbb{N}$, $\varepsilon \in (0,1)$, $j \in \{0,1,\dots,N\}$, $x \in [0,1]$ that

$$\begin{aligned}
|[R_\varrho(\tau_{f,\varepsilon,N,j})](x)| &\le |[R_\varrho(\tau_{f,\varepsilon,N,j})](x) - T_{f,N,j}(x)| + |T_{f,N,j}(x) - f(x)| + |f(x)| \\
&\le \frac{\varepsilon}{4(N+1)} + \tfrac{1}{n!}x^n \sup_{t\in[0,1]}|f^{(n)}(t)| + \sup_{t\in[0,1]}|f(t)| \le 3.
\end{aligned} \tag{A.80}$$

Combining this with Lemma 5.3, Lemma 5.4, (A.60), (C), (A.78), and (A.79) establishes for every $f \in B^n_1$, $N \in \mathbb{N}$, $\varepsilon \in (0,1)$, $j \in \{0,1,\dots,N\}$, $x \in [0,1]$ the bound

$$\begin{aligned}
&|h_{N,j}(x)T_{f,N,j}(x) - [R_\varrho(\psi_{f,\varepsilon,N,j})](x,x)| \\
&\le |h_{N,j}(x)T_{f,N,j}(x) - [R_\varrho(\chi_{N,j})](x)[R_\varrho(\tau_{N,j})](x)| \\
&\quad + \left|[R_\varrho(\chi_{N,j})](x)[R_\varrho(\tau_{N,j})](x) - [R_\varrho(\Pi^2_{\varepsilon/8} \circ \mathcal{P}(\chi_{N,j},\tau_{f,\varepsilon,N,j}))](x,x)\right| \\
&\le |h_{N,j}(x)T_{f,N,j}(x) - [R_\varrho(\tau_{N,j})](x)| \\
&\quad + \left|[R_\varrho(\chi_{N,j})](x)[R_\varrho(\tau_{N,j})](x) - [R_\varrho(\Pi^2_{\varepsilon/8})]([R_\varrho(\chi_{N,j}](x),[R_\varrho(\tau_{f,\varepsilon,N,j})](x))\right| \\
&\le \tfrac{\varepsilon}{8} + \tfrac{\varepsilon}{8} = \tfrac{\varepsilon}{4}.
\end{aligned} \tag{A.81}$$

Furthermore, note that for every $N \in \mathbb{N}$, $j \in \{0, 1, \ldots, N\}$, $x \notin [\frac{j-1}{N}, \frac{j+1}{N}]$ it holds that $h_{N,j}(x) = \chi_{N,j}(x) = 0$. Thus (D) ensures that for every $f \in B_1^n$, $N \in \mathbb{N}$, $\varepsilon \in (0,1)$, $j \in \{0, 1, \ldots, N\}$, $x \in [0,1]$ it holds

$$|h_{N,j}(x) T_{f,N,j}(x) - [R_\varrho(\psi_{f,\varepsilon,N,j})](x,x)| = 0. \tag{A.82}$$

This, Lemma 5.3, Lemma 5.4, (A.62), (A.74), and (A.81) imply that for every $f \in B_1^n$, $N \in \mathbb{N}$, $\varepsilon \in (0,1)$, $x \in [0,1]$ it holds

$$\begin{aligned} |f_N(x) - [R_\varrho(\varphi_{f,\varepsilon,N})](x)| &= \left| \sum_{j=0}^{N} h_{N,j}(x) T_{f,N,j}(x) - \sum_{j=0}^{N} [R_\varrho(\psi_{f,\varepsilon,N,j})](x,x) \right| \\ &\leq 2 \max_{j \in \{0,1,\ldots,N\}} |h_{N,j}(x) T_{f,N,j}(x) - [R_\varrho(\psi_{f,\varepsilon,N,j})](x,x)| \\ &\leq \tfrac{\varepsilon}{2}. \end{aligned} \tag{A.83}$$

Combining this with (A.66) establishes that for every $f \in B_1^n$, $N \in \mathbb{N}$, $\varepsilon \in (0,1)$, $x \in [0,1]$ it holds

$$|f(x) - [R_\varrho(\varphi_{f,\varepsilon,N})](x)| \leq |f(x) - f_N(x)| + |f_N(x) - [R_\varrho(\varphi_{f,\varepsilon,N})]| \leq \tfrac{1}{n!} N^{-n} + \tfrac{\varepsilon}{2}. \tag{A.84}$$

Let $N_\varepsilon \in \mathbb{N}$ satisfy for every $\varepsilon \in (0,\infty)$

$$N_\varepsilon = \left\lceil \left[\tfrac{2}{n!\varepsilon}\right]^{1/n} \right\rceil, \tag{A.85}$$

let $\theta \in \mathcal{N}_1^{1,1}$ be given by $\theta = (0,0)$, and let $(\Phi_{f,\varepsilon})_{f \in B_1^n, \varepsilon \in (0,\infty)} \subseteq \mathfrak{N}$ be the neural networks given by

$$\Phi_{f,\varepsilon} = \begin{cases} \varphi_{f,\varepsilon,N_\varepsilon} & : \varepsilon < 1 \\ \theta & : \varepsilon \geq 1 \end{cases}. \tag{A.86}$$

Oberve that (A.84) implies that for every $f \in B_1^n$, $\varepsilon \in (0,1)$, $x \in [0,1]$

$$|f(x) - [R_\varrho(\Phi_{f,\varepsilon})](x)| = |f(x) - [R_\varrho(\varphi_{f,\varepsilon,N_\varepsilon})](x)| \leq \tfrac{1}{n!} N_\varepsilon^{-n} + \tfrac{\varepsilon}{2} \leq \tfrac{1}{n!} \left[\tfrac{n!\varepsilon}{2}\right] + \tfrac{\varepsilon}{2} = \varepsilon. \tag{A.87}$$

Moreover that for every $f \in B_1^n$, $\varepsilon \in [1,\infty)$, $x \in [0,1]$ it holds

$$|f(x) - [R_\varrho(\Phi_{f,\varepsilon})](x)| = |f(x) - [R_\varrho(\theta)](x)| = |f(x)| \leq 1 \leq \varepsilon. \tag{A.88}$$

This and (A.87) establish that the neural networks $(\Phi_{f,\varepsilon})_{f \in B_1^n, \varepsilon \in (0,\infty)}$ satisfy (iii).

Next, Lemma 5.3, Lemma 5.4, (A), (A.67), and (A.68) imply for every $N \in \mathbb{N}$, $\varepsilon \in (0,\infty)$, $j \in \{0, 1, \ldots, N\}$, $k \in \{1, 2, \ldots, n-1\}$

$$\mathcal{L}(\xi_{\varepsilon,N,j}^k) \leq \max\{1, \mathcal{L}(\Pi_{\varepsilon/8e}^k) + \mathcal{L}(\nabla_{N,j,k})\} \leq C \ln(k) \left(|\ln(\tfrac{\varepsilon}{8e})| + k\ln(3) + \ln(k)\right) + 1. \tag{A.89}$$

Combining this with Lemma 5.3, Lemma 5.4, (A.67), (A.69), (A.70) shows for every $f \in B_1^n$, $N \in \mathbb{N}$, $\varepsilon \in (0,\infty)$, $j \in \{0, 1, \ldots, N\}$ the bound

$$\begin{aligned} \mathcal{L}(\tau_{f,\varepsilon,N,j}) &\leq \mathcal{L}(\Sigma_{f,N,j}) + \left[\max_{k \in \{1,2,\ldots,n-1\}} \mathcal{L}(\xi_{\varepsilon,N,j}^k)\right] + \mathcal{L}(\nabla_{1,0,n-1}) \\ &\leq 3 + C\ln(n) \left(|\ln(\tfrac{\varepsilon}{8e})| + n\ln(3) + \ln(n)\right). \end{aligned} \tag{A.90}$$

This, Lemma 5.3, Lemma 5.4, (A), (A.71), (A.72), (A.74), and (A.67) ensure for every $f \in B_1^n$, $N \in \mathbb{N}$, $\varepsilon \in (0,\infty)$ it holds

$$\begin{aligned} \mathcal{L}(\varphi_{f,\varepsilon,N}) &\leq \mathcal{L}(\lambda_N) + \left[\max_{j \in \{0,1,\ldots,N\}} \mathcal{L}(\psi_{f,\varepsilon,N,j})\right] + \mathcal{L}(\nabla_{1,0,2N+2}) \\ &\leq 2 + \left[\max_{j \in \{0,1,\ldots,N\}} \mathcal{L}(\Pi_{\varepsilon/8}^2 \circ \mathcal{P}(\chi_{N,j}, \tau_{f,\varepsilon,N,j}))\right] \\ &\leq 2 + \left[C\ln(2)\left(|\ln(\tfrac{\varepsilon}{8})| + 2\ln(3) + \ln(2)\right) + \max\{3, \mathcal{L}(\tau_{f,\varepsilon,N,j})\}\right] \\ &\leq 5 + C\ln(2)\left(|\ln(\tfrac{\varepsilon}{8})| + \ln(18)\right) + C\ln(n)\left(|\ln(\tfrac{\varepsilon}{8e})| + n\ln(3) + \ln(n)\right) \\ &\leq 5 + C\ln(2)\left(|\ln(\varepsilon)| + |\ln(8)| + \ln(18)\right) \\ &\quad + C\ln(n)\left(|\ln(\varepsilon)| + |\ln(8e)| + n\ln(3) + \ln(n)\right) \\ &= C\ln(2n)\,|\ln(\varepsilon)| + C(\ln(2)\ln(144) + \ln(n)(\ln(3)n + \ln(n) + |\ln(8e)|)) + 5. \end{aligned} \tag{A.91}$$

With the constant C from (A.91), define the term T_1 by

$$T_1 = C(\ln(2)\ln(144) + \ln(n)(\ln(3)n + \ln(n) + |\ln(8e)|)) + 5. \tag{A.92}$$

Observe that (A.91) implies for every $f \in B_1^n$, $\varepsilon \in (0,1)$

$$\mathcal{L}(\Phi_{f,\varepsilon}) = \mathcal{L}(\varphi_{f,\varepsilon,N_\varepsilon}) = C\ln(2n)\,|\ln(\varepsilon)| + T_1. \tag{A.93}$$

Hence we obtain

$$\sup_{f\in B_1^n,\varepsilon\in(0,e^{-r}]} \left[\tfrac{\mathcal{L}(\Phi_{f,\varepsilon})}{\max\{r,|\ln(\varepsilon)|\}}\right] \le \sup_{f\in B_1^n,\varepsilon\in(0,e^{-r}]} \left[\tfrac{C\ln(2n)|\ln(\varepsilon)|+T_1}{|\ln(\varepsilon)|}\right] \le C\ln(2n) + \tfrac{T_1}{r} < \infty. \tag{A.94}$$

In addition, note that (A.93) ensures that

$$\sup_{f\in B_1^n,\varepsilon\in(e^{-r},1)} \left[\tfrac{\mathcal{L}(\Phi_{f,\varepsilon})}{\max\{r,|\ln(\varepsilon)|\}}\right] \le \sup_{f\in B_1^n,\varepsilon\in(e^{-r},1)} \left[\frac{C\ln(2n)\,|\ln(\varepsilon)| + T_1}{r}\right] \le C\ln(2n) + \tfrac{T_1}{r} < \infty. \tag{A.95}$$

Furthermore

$$\sup_{f\in B_1^n,\varepsilon\in[1,\infty)} \left[\frac{\mathcal{L}(\Phi_{f,\varepsilon})}{\max\{r,|\ln(\varepsilon)|\}}\right] = \sup_{f\in B_1^n,\varepsilon\in[1,\infty)} \left[\frac{1}{\max\{r,|\ln(\varepsilon)|\}}\right] < \infty. \tag{A.96}$$

This, (A.94), and (A.95) establish that the neural networks $(\Phi_{f,\varepsilon})_{\varepsilon\in(0,\infty)}$ satisfy (i). Next, Lemma 5.3, (B), (A.67), and (A.68) imply for every $N \in \mathbb{N}$, $\varepsilon \in (0,\infty)$, $j \in \{0,1,\dots,N\}$, $k \in \{1,2,\dots,n-1\}$

$$\mathcal{M}(\xi^k_{\varepsilon,N,j}) \le \max\{1, 2(\mathcal{M}(\Pi^k_{\varepsilon/8e}) + \mathcal{M}(\nabla_{N,j,k}))\} \le 2(Ck\left(\left|\ln(\tfrac{\varepsilon}{8e})\right| + k\ln(3) + \ln(k)\right) + 1) \tag{A.97}$$

Combining this with Lemma 5.3, Lemma 5.4, (A.67), (A.69), and (A.70) shows for every $f \in B_1^n$, $N \in \mathbb{N}$, $\varepsilon \in (0,\infty)$, $j \in \{0,1,\dots,N\}$ it holds

$$\begin{aligned}
\mathcal{M}(\tau_{f,\varepsilon,N,j}) &\le 2\left(\mathcal{M}(\Sigma_{f,N,j}) + 2\left(\mathcal{M}(\mathcal{P}(\xi^{n-1}_{\varepsilon,N,j},\dots,\xi^1_{\varepsilon,N,j})) + \mathcal{L}(\nabla_{1,0,n-1})\right)\right)\\
&\le 2n + 4\left(2\left[\sum_{k=1}^{n-1}\mathcal{M}(\xi^k_{\varepsilon,N,j})\right] + 4(n-1)\max_{k\in\{1,2,\dots,n-1\}}\mathcal{L}(\xi^k_{\varepsilon,N,j})\right) + 8(n-1)\\
&\le 10n + 8(n-1)(2Cn\left(\ln(\tfrac{\varepsilon}{(8e)})| + n\ln(3) + \ln(n)\right) + 2)\\
&\quad + 16(n-1)(C\ln(n)\left(|\ln(\tfrac{\varepsilon}{8e})| + n\ln(3) + \ln(n)\right) + 1)\\
&\le 32n^2C\left(|\ln(\tfrac{\varepsilon}{8e})| + n\ln(3) + \ln(n)\right) + 42n.
\end{aligned} \tag{A.98}$$

Let the term T_2 be given by

$$T_2 = 128\left(C + 32n^2C + C\ln(n)\right), \tag{A.99}$$

and let the term T_3 be given by

$$T_3 = 1556 + 128(C\ln(144) + 64n^2C(n\ln(3) + \ln(n)) + 42n. \tag{A.100}$$

This, Lemma 5.3, Lemma 5.4, (B), (A.67), (A.71), (A.72), (A.74), and the fact that for every $\psi \in \mathfrak{N}$ with $\min_{l\in\{1,2,\dots,\mathcal{L}(\psi)\}} \mathcal{M}_l(\psi) > 0$ it holds that $\mathcal{L}(\psi) \le \mathcal{M}(\psi)$ ensure that for every $f \in B_1^n$, $N \in \mathbb{N}$, $\varepsilon \in (0,\infty)$ it holds

$$
\begin{aligned}
&\mathcal{M}(\varphi_{f,\varepsilon,N}) \\
&\leq 2\left(\mathcal{M}(\lambda_N) + 2\left[\mathcal{M}(\mathcal{P}(\psi_{f,\varepsilon,N,1}, \psi_{f,\varepsilon,N,2}, \ldots, \psi_{f,\varepsilon,N,N})) + \mathcal{M}(\nabla_{1,0,2N+2})\right]\right) \\
&\leq 2(N+1) + 8\left[\sum_{j=0}^{N} \mathcal{M}(\psi_{f,\varepsilon,N,j})\right] + 16(N+1)\left[\max_{j\in\{0,1,\ldots,N\}} \mathcal{L}(\psi_{f,\varepsilon,N,j})\right] + 8(N+1) \\
&\leq 20N + 32(N+1) \max_{j\in\{1,2,\ldots,N\}} \mathcal{M}(\psi_{f,\varepsilon,N,j}) \\
&\leq 20N + 64N\left(\mathcal{M}(\Pi^2_{\varepsilon/8}) + \mathcal{M}(\mathcal{P}(\chi_{N,N}, \tau_{f,\varepsilon,N,N}))\right) \\
&\leq 20N + 128NC\left(\left|\ln(\tfrac{\varepsilon}{8})\right| + 2\ln(3) + \ln(2)\right) \\
&\quad + 64N\left(2\mathcal{M}(\chi_{N,N}) + 2\mathcal{M}(\tau_{f,\varepsilon,N,N}) + 4\max\{\mathcal{L}(\chi_{N,N}), \mathcal{L}(\tau_{f,\varepsilon,N,N})\}\right) \\
&\leq 20N + 128NC\left(\left|\ln(\tfrac{\varepsilon}{8})\right| + \ln(18)\right) + 1152N \\
&\quad + 128N\left(32n^2C\left(|\ln(\tfrac{\varepsilon}{8e})| + n\ln(3) + \ln(n)\right) + 42n\right) \\
&\quad + 128N\left(3 + C\ln(n)\left(|\ln(\tfrac{\varepsilon}{8e})| + n\ln(3) + \ln(n)\right)\right) \\
&= 128\left(C + 32n^2C + C\ln(n)\right)N|\ln(\varepsilon)| \\
&\quad + \left(1556 + 128(C\ln(144) + 64n^2C(n\ln(3) + \ln(n)) + 42n\right)N \\
&= T_2N|\ln(\varepsilon)| + T_3N.
\end{aligned}
\tag{A.101}
$$

Combining this with Lemma A.2 demonstrates that for every $f \in B_1^n$, $\varepsilon \in (0, \exp(-2n^2)]$ it holds

$$
\begin{aligned}
\mathcal{M}(\Phi_{f,\varepsilon}) = \mathcal{M}(\varphi_{f,\varepsilon,N_\varepsilon}) &\leq T_2N_\varepsilon|\ln(\varepsilon)| + T_3N_\varepsilon \\
&= T_2\left\lceil \left[\tfrac{2}{n!\varepsilon}\right]^{1/n}\right\rceil |\ln(\varepsilon)| + T_3\left\lceil \left[\tfrac{2}{n!\varepsilon}\right]^{1/n}\right\rceil \\
&\leq 3T_2\varepsilon^{-\frac{1}{n}}|\ln(\varepsilon)| + 3T_3\varepsilon^{-\frac{1}{n}} \\
&\leq 3T_2\varepsilon^{-\frac{1}{n}}\max\{r, |\ln(\varepsilon)|\} + 3T_3\varepsilon^{-\frac{1}{n}}.
\end{aligned}
\tag{A.102}
$$

Hence we obtain

$$
\sup_{f\in B_1^n, \varepsilon\in(0,\exp(-2n^2))}\left[\frac{\mathcal{M}(\Phi_{f,\varepsilon})}{\varepsilon^{-\frac{1}{n}}\max\{r, |\ln(\varepsilon)|\}}\right] \leq 3T_2 + 3T_3\frac{1}{\max\{r, 2n^2\}} < \infty. \tag{A.103}
$$

Combining (A.102) with the fact that continuous function are bounded on compact sets ensures

$$
\begin{aligned}
&\sup_{f\in B_1^n, \varepsilon\in[\exp(-2n^2),1]}\left[\frac{\mathcal{M}(\Phi_{f,\varepsilon})}{\varepsilon^{-\frac{1}{n}}\max\{r, |\ln(\varepsilon)|\}}\right] \\
&\leq \sup_{f\in B_1^n, \varepsilon\in[\exp(-2n^2),1]}\left[\frac{T_2N(|\ln(\varepsilon)| + |\ln(N)|) + T_3N}{\varepsilon^{-\frac{1}{n}}\max\{r, |\ln(\varepsilon)|\}}\right] < \infty.
\end{aligned}
\tag{A.104}
$$

In addition note

$$
\sup_{f\in B_1^n, \varepsilon\in(1,\infty)}\left[\frac{\mathcal{M}(\Phi_{f,\varepsilon})}{\varepsilon^{-\frac{1}{n}}\max\{r, |\ln(\varepsilon)|\}}\right] = \sup_{f\in B_1^n, \varepsilon\in(1,\infty)}\left[\frac{\mathcal{M}(\theta)}{\varepsilon^{-\frac{1}{n}}\max\{r, |\ln(\varepsilon)|\}}\right] \tag{A.105}
$$

$$
= \sup_{f\in B_1^n, \varepsilon\in(1,\infty)}\left[\frac{0}{\varepsilon^{-\frac{1}{n}}\max\{r, |\ln(\varepsilon)|\}}\right] = 0 < \infty. \tag{A.106}
$$

This, (A.103), and (A.104) establish that the neural networks $(\Phi_{f,\varepsilon})_{f\in B_1^n, \varepsilon\in(0,\infty)}$ satisfy (ii). The proof of Theorem 6.5 is completed. □

A.6 Proof of Corollary 6.6

Proof of Corollary 6.6. Throughout this proof assume Setting 5.2, let $c_{a,b} \in \mathbb{R}$, $[a,b] \subseteq \mathbb{R}_+$, be the real numbers given by $c_{a,b} = \min\{1, (b-a)^{-n}\}$, let $\lambda_{a,b} \in \mathcal{N}_1^{1,1}$, $[a,b] \subseteq \mathbb{R}_+$, be the neural networks given by $\lambda_{a,b} = (\frac{1}{b-a}, -\frac{a}{b-a})$, let $\alpha_f \in \mathcal{N}_1^{1,1}$, $f \in \mathcal{C}^n$ be the neural networks given by $\alpha_f = (\frac{1}{c} \|f\|_{n,\infty}, 0)$, let $L_{a,b} \colon [0,1] \to [a,b]$, $[a,b] \subseteq \mathbb{R}_+$ be the functions which satisfy for every $[a,b] \subseteq \mathbb{R}_+$, $t \in [0,1]$

$$L_{a,b}(t) = (b-a)t + a, \tag{A.107}$$

and for every $f \in \mathcal{C}^n$ let $f_* \in C^n([0,1], \mathbb{R})$ be the function which satisfies for every $t \in [0,1]$

$$f_*(t) = \|f\|_{n,\infty}^{-1} c_{a,b}(f(L_{a,b}(t))). \tag{A.108}$$

We claim that for every $[a,b] \subseteq \mathbb{R}_+$, $f \in C^n([a,b], \mathbb{R})$, $m \in \{1, 2, \ldots, n\}$, $t \in [0,1]$ it holds

$$f_*^{(m)}(t) = \|f\|_{n,\infty}^{-1} c_{a,b}(b-a)^m [f^{(m)}(L_{a,b}(t))]. \tag{A.109}$$

We now prove (A.109) by induction on $m \in \{1, 2, \ldots, n\}$. For the base case $m = 1$, the chain rule implies for every $[a,b] \subseteq \mathbb{R}_+$, $f \in C^n([a,b], \mathbb{R})$, $t \in [0,1]$

$$\begin{aligned} f_*'(t) &= \tfrac{\mathrm{d}}{\mathrm{d}t} \Big[\|f\|_{n,\infty}^{-1} c_{a,b} f(L_{a,b}(t)) \Big] = \|f\|_{n,\infty}^{-1} c_{a,b} \left[f'(L_{a,b}(t)) L_{a,b}'(t) \right] \\ &= \|f\|_{n,\infty}^{-1} c_{a,b} \left[f'(L_{a,b}(t))(b-a) \right] = \|f\|_{n,\infty}^{-1} c_{a,b}(b-a)[f'(L_{a,b}(t))]. \end{aligned} \tag{A.110}$$

This establishes (A.109) in the base case $m = 1$.

For the induction step $\{1, 2, \ldots, n-1\} \ni m \to m+1 \in \{2, 3, \ldots, n\}$ observe that the chain rule ensures for every $[a,b] \subseteq \mathbb{R}_+$, $f \in C^n([a,b], \mathbb{R})$, $m \in \mathbb{N}$, $t \in [0,1]$

$$\begin{aligned} \tfrac{\mathrm{d}}{\mathrm{d}t} \Big[\|f\|_{n,\infty}^{-1} c_{a,b}(b-a)^m [f^{(m)}(L_{a,b}(t))] \Big] &= \|f\|_{n,\infty}^{-1} c_{a,b}(b-a)^m [f^{(m+1)}(L_{a,b}(t)) L_{a,b}'(t)] \\ &= \|f\|_{n,\infty}^{-1} c_{a,b}(b-a)^{m+1} [f^{(m+1)}(L_{a,b}(t))]. \end{aligned} \tag{A.111}$$

Induction thus establishes (A.109).

In addition, for every $[a,b] \subseteq \mathbb{R}_+$, $k \in \{0, 1, \ldots, n\}$

$$c_{a,b}(b-a)^k = \min\{1, (b-a)^{-n}\}(b-a)^k = \min\{(b-a)^k, (b-a)^{-n+k}\} \leq 1. \tag{A.112}$$

Combining this with (6.30), (A.107), and (A.109) ensures for every $[a,b] \subseteq \mathbb{R}_+$, $f \in C^n([a,b], \mathbb{R})$

$$\begin{aligned} \max_{k \in \{0,1,\ldots,n\}} \left[\sup_{t \in [0,1]} \left| f_*^{(k)}(t) \right| \right] &= \max_{k \in \{0,1,\ldots,n\}} \left[\sup_{t \in [a,b]} \left| \|f\|_{n,\infty}^{-1} c_{a,b}(b-a)^k [f^{(k)}(t)] \right| \right] \\ &\leq \|f\|_{n,\infty}^{-1} \max_{k \in \{0,1,\ldots,n\}} \left[\sup_{t \in [a,b]} \left| f^{(k)}(t) \right| \right] = 1. \end{aligned} \tag{A.113}$$

Theorem 6.5 therefore establishes that there exist neural networks $(\Phi_{g,\eta})_{g \in B_1^n, \eta \in (0,\infty)} \subseteq \mathfrak{N}$ which satisfy

(a) $\displaystyle\sup_{g \in B_1^n, \eta \in (0,\infty)} \left[\frac{\mathcal{L}(\Phi_{g,\eta})}{\max\{r, |\ln(\eta)|\}} \right] < \infty$,

(b) $\displaystyle\sup_{g \in B_1^n, \eta \in (0,\infty)} \left[\frac{\mathcal{M}(\Phi_{g,\eta})}{\eta^{-\frac{1}{n}} \max\{r, |\ln(\eta)|\}} \right] < \infty$, and

(c) for every $g \in B_1^n$, $\eta \in (0,\infty)$ that

$$\sup_{t \in [0,1]} |g(t) - [R_\varrho(\Phi_{g,\eta})](t)| \leq \eta. \tag{A.114}$$

Let $(\Phi_{f,\varepsilon})_{f\in\mathcal{C}^n,\varepsilon\in(0,\infty)} \subseteq \mathfrak{N}$ denote neural networks which satisfy for every $[a,b] \subseteq \mathbb{R}_+$, $f \in C^n([a,b],\mathbb{R})$, $\varepsilon \in (0,\infty)$

$$\Phi_{f,\varepsilon} = \alpha_f \circ \varphi_{f_*,\frac{c_{a,b}\varepsilon}{\|f\|_{n,\infty}}} \circ \lambda_{a,b}. \tag{A.115}$$

Observe that for every $[a,b] \subseteq \mathbb{R}_+$, $f \in C^n([a,b],\mathbb{R})$, $t \in [0,1]$ it holds

$$[R_\varrho(\lambda_{a,b})](t) = \left[\tfrac{1}{(b-a)}\right] t - \tfrac{a}{(b-a)} = L_{a,b}^{-1}(t) \qquad \text{and} \qquad [R_\varrho(\alpha_f)](t) = \tfrac{\|f\|_{n,\infty}}{c_{a,b}} t. \tag{A.116}$$

Lemma 5.3 therefore demonstrates for every $[a,b] \subseteq \mathbb{R}_+$, $f \in C^n([a,b],\mathbb{R})$, $\varepsilon \in (0,\infty)$, $t \in [0,1]$ it holds

$$\begin{aligned}
[R_\varrho(\Phi_{f,\varepsilon})](t) &= [R_\varrho(\alpha_f \circ \varphi_{f_*,\frac{c_{a,b}\varepsilon}{\|f\|_{n,\infty}}} \circ \lambda_{a,b})](t) \\
&= [R_\varrho(\alpha_f) \circ R_\varrho(\varphi_{f_*,\frac{c_{a,b}\varepsilon}{\|f\|_{n,\infty}}}) \circ R_\varrho(\lambda_{a,b})](t) \\
&= \tfrac{\|f\|_{n,\infty}}{c_{a,b}} [R_\varrho(\varphi_{f_*,\frac{c_{a,b}\varepsilon}{\|f\|_{n,\infty}}})](L_{a,b}^{-1}(t)).
\end{aligned} \tag{A.117}$$

Moreover, note (A.108) ensures that for every $[a,b] \subseteq \mathbb{R}_+$, $f \in C^n([a,b],\mathbb{R})$, $t \in [a,b]$ it holds

$$f(t) = \tfrac{\|f\|_{n,\infty}}{c_{a,b}} f_*(L_{a,b}^{-1}(t)). \tag{A.118}$$

Combining (c), (A.115), and (A.117) implies for every $[a,b] \subseteq \mathbb{R}_+$, $f \in C^n([a,b],\mathbb{R})$, $\varepsilon \in (0,\infty)$

$$\begin{aligned}
\sup_{t\in[a,b]} |f(t) - [R_\varrho(\Phi_{f,\varepsilon})](t)| &= \sup_{t\in[a,b]} \left| \tfrac{\|f\|_{n,\infty}}{c_{a,b}} f_*(L_{a,b}^{-1}(t)) - \tfrac{\|f\|_{n,\infty}}{c_{a,b}} [R_\varrho(\varphi_{f_*,\frac{c_{a,b}\varepsilon}{\|f\|_{n,\infty}}})](L_{a,b}^{-1}(t)) \right| \\
&= \tfrac{\|f\|_{n,\infty}}{c_{a,b}} \left[\sup_{t\in[0,1]} \left| f_*(t) - [R_\varrho(\varphi_{f_*,\frac{c_{a,b}\varepsilon}{\|f\|_{n,\infty}}})](t) \right| \right] \leq \tfrac{\|f\|_{n,\infty}}{c_{a,b}} \frac{c_{a,b}\varepsilon}{\|f\|_{n,\infty}} = \varepsilon.
\end{aligned} \tag{A.119}$$

This establishes that the neural networks $(\Phi_{f,\varepsilon})_{f\in\mathcal{C}^n,\varepsilon\in(0,\infty)}$ satisfy (iii). Furthermore, Lemma 5.3 ensures for every $[a,b] \subseteq \mathbb{R}_+$, $f \in C^n([a,b],\mathbb{R})$, $\varepsilon \in (0,\infty)$ holds

$$\mathcal{L}(\Phi_{f,\varepsilon}) = \mathcal{L}(\alpha_f \circ \varphi_{f_*,\frac{c_{a,b}\varepsilon}{\|f\|_{n,\infty}}} \circ \lambda_{a,b}) = \mathcal{L}(\alpha_f) + \mathcal{L}(\varphi_{f_*,\frac{c_{a,b}\varepsilon}{\|f\|_{n,\infty}}}) + \mathcal{L}(\lambda_{a,b}) = \mathcal{L}(\varphi_{f_*,\frac{c_{a,b}\varepsilon}{\|f\|_{n,\infty}}}) + 2. \tag{A.120}$$

In addition, for every $[a,b] \subseteq \mathbb{R}_+$, $f \in C^n([a,b],\mathbb{R})$, $\varepsilon \in (0,\infty)$ holds

$$\begin{aligned}
\max\{r, |\ln(\tfrac{c_{a,b}\varepsilon}{\|f\|_{n,\infty}})|\} &= \max\{r, |\ln(\tfrac{\min\{1,(b-a)^{-n}\}\varepsilon}{\|f\|_{n,\infty}})|\} = \max\{r, |\ln(\tfrac{\varepsilon}{(\max\{1,(b-a)\})^n \|f\|_{n,\infty}})|\} \\
&\leq n \max\{r, |\ln(\tfrac{\varepsilon}{(\max\{1,(b-a)\})\|f\|_{n,\infty}})|\}.
\end{aligned} \tag{A.121}$$

Combining this with (a) and (A.120) implies that

$$\begin{aligned}
\sup_{f\in\mathcal{C}^n,\varepsilon\in(0,\infty)} \left[\frac{\mathcal{L}(\Phi_{f,\varepsilon})}{\max\{r, |\ln(\frac{\varepsilon}{\max\{1,b-a\}\|f\|_{n,\infty}})|\}} \right] &\leq n \sup_{f\in\mathcal{C}^n,\varepsilon\in(0,\infty)} \left[\frac{\mathcal{L}(\varphi_{f_*,\frac{c_{a,b}\varepsilon}{\|f\|_{n,\infty}}}) + 2}{\max\{r, |\ln(\frac{c_{a,b}\varepsilon}{\|f\|_{n,\infty}})|\}} \right] \\
&= n \sup_{g\in B_1^n,\eta\in(0,\infty)} \left[\frac{\mathcal{L}(\Phi_{g,\eta}) + 2}{\max\{r, |\ln(\eta)|\}} \right] < \infty.
\end{aligned} \tag{A.122}$$

This establishes that the neural networks $(\Phi_{f,\varepsilon})_{f\in\mathcal{C}^n,\varepsilon\in(0,\infty)}$ satisfy (i). Next, Lemma 5.3 implies that for every $[a,b] \subseteq \mathbb{R}_+$, $f \in C^n([a,b],\mathbb{R})$, $\varepsilon \in (0,\infty)$

$$\mathcal{M}(\Phi_{f,\varepsilon}) = \mathcal{M}(\alpha_f \circ \varphi_{f_*,\frac{c_{a,b}\varepsilon}{\|f\|_{n,\infty}}} \circ \lambda_{a,b}) = \mathcal{M}(\alpha_f) + \mathcal{M}(\varphi_{f_*,\frac{c_{a,b}\varepsilon}{\|f\|_{n,\infty}}}) + \mathcal{M}(\lambda_{a,b}) = \mathcal{M}(\varphi_{f_*,\frac{c_{a,b}\varepsilon}{\|f\|_{n,\infty}}}) + 3. \tag{A.123}$$

In addition, note that (A.121) shows for every $[a,b] \subseteq \mathbb{R}_+$, $f \in C^n([a,b],\mathbb{R})$, $\varepsilon \in (0,\infty)$

$$\left[\tfrac{c_{a,b}\varepsilon}{\|f\|_{n,\infty}}\right]^{-\frac{1}{n}} \max\{r, |\ln(\tfrac{c_{a,b}\varepsilon}{\|f\|_{n,\infty}})|\} n \leq \max\{1, b-a\} \, \|f\|_{n,\infty}^{\frac{1}{n}} \, \varepsilon^{-\frac{1}{n}} \max\{r, |\ln(\tfrac{\varepsilon}{\max\{1,b-a\}\|f\|_{n,\infty}})|\}. \tag{A.124}$$

Combining this with (b) and (A.115) therefore ensures

$$\begin{aligned}
&\sup_{f\in\mathcal{C}^n, \varepsilon\in(0,\infty)} \left[\frac{\mathcal{M}(\Phi_{f,\varepsilon})}{\max\{1,b-a\}\,\|f\|_{n,\infty}^{\frac{1}{n}}\,\varepsilon^{-\frac{1}{n}} \max\{r, |\ln(\frac{\varepsilon}{\max\{1,b-a\}\|f\|_{n,\infty}})|\}}\right] \\
&\leq n \sup_{f\in\mathcal{C}^n, \varepsilon\in(0,\infty)} \left[\frac{\mathcal{M}(\varphi_{f_*, \frac{c_{a,b}\varepsilon}{\|f\|_{n,\infty}}}) + 3}{\left[\frac{c_{a,b}\varepsilon}{\|f\|_{n,\infty}}\right]^{-\frac{1}{n}} \max\{r, |\ln(\frac{c_{a,b}\varepsilon}{\|f\|_{n,\infty}})|\}}\right] \\
&\leq n \sup_{g\in B_1^n, \eta\in(0,\infty)} \left[\frac{\mathcal{M}(\Phi_{g,\eta}) + 3}{\eta^{-\frac{1}{n}} \max\{r, |\ln(\eta)|\}}\right] < \infty.
\end{aligned} \tag{A.125}$$

This establishes that the neural networks $(\Phi_{f,\varepsilon})_{f\in\mathcal{C}^n, \varepsilon\in(0,\infty)}$ satisfy (ii) and completes the proof. □